arXiv:2602.07944v1 [math.ST] 8 Feb 2026

GIG方差混合线性潜变量模型的吉布斯采样器的几何遍历性

作者：ELSIDDIG AWADELKARIM[1,*,d]、DAVID BOLIN[1,a]、XIAOTIAN JIN[1,†,b]、ALEXANDRE B. SIMAS[1,c] 和 JONAS WALLIN[2,e]

[1]*CEMSE*，阿卜杜拉国王科技大学，沙特阿拉伯，[a]*david.bolin@kaust.edu.sa*；[b]*xiaotian.jin@kaust.edu.sa*；[c]*alexandre.simas@kaust.edu.sa*；[d]*elsiddig.awadelkarim@kaust.edu.sa*

[2]隆德大学统计学系，瑞典，[e]*jonas.wallin@stat.lu.se*

我们研究线性潜在非高斯模型（LLnGMs）中吉布斯采样器的几何遍历性。这类层次模型通过广义逆高斯分布（GIG）方差混合增广保留了条件高斯结构。我们为混合变量上的边缘链提出了两种互补的几何遍历性证明途径。首先，我们证明相应的马尔可夫算子是迹类，因此在GIG参数空间的大部分区域内存在谱隙。其次，对于其余的边界和重尾情形，我们通过漂移与小化证明几何遍历性，但需满足一个明确的零空间小性条件，该条件量化了漂移与观测算子零空间之间的相互作用。综合而言，这些结果覆盖了整个GIG参数空间，包括正态-逆高斯、广义非对称拉普拉斯和Student-t 特殊情形。该链的几何遍历性为基于吉布斯采样的随机梯度估计量用于最大似然估计提供了理论基础；我们还给出了相应条件，使所需的可积性检验一目了然。数值实验展示了理论结果，对比了不同参数区域下的混合效率，并探究了零空间小性常数的作用。

1. 引言。

1.1. 马尔可夫链的几何遍历性。 马尔可夫链蒙特卡罗（MCMC）方法是一类旨在近似概率分布的采样算法。这些方法已成为几乎所有科学学科中统计计算的基础工具。对于马尔可夫链，遍历性确保其收敛到唯一的平稳分布，且该分布与初始分布无关。这一性质至关重要，因为它保证沿单条轨迹计算的时间平均值会收敛到平稳分布下的总体平均值。几何遍历性是遍历性的强化形式，它量化了收敛速度，并保证链以指数（几何）速度收敛到其平稳分布。

若马尔可夫链 $\{X_t\}_{t\geq 0}$ 的状态空间为 $\mathcal{X}$，平稳分布为 π，则称其具有几何遍历性；如果存在常数 $C>0$、$\gamma\in(0,\,1)$，使得对于所有概率测度 $\mu\in\mathcal{L}_2(\pi)$，

$$\|\mu P^t-\pi\|_{\mathrm{TV}}\leq C\gamma^t,$$

其中，P是与链 $t\in\mathbb{N}$ 相关联的马尔可夫算子，$\|\cdot\|_{\mathrm{TV}}$ 表示全变差距离。这里，$\mathcal{L}_2(\pi)$ 是 $\mathcal{X}$ 上的符号测度空间，这些测度相对于 $\mathcal{X}$ 绝对连续，

*作者按字母顺序排列。

†通讯作者。*MSC2020*主题分类：主要分类：60J10、65C40；次要分类：37A25、60J22。关键词和短语：几何遍历性、迹类、吉布斯采样器、非高斯模型、正态方差-均值混合分布。

且其拉东–尼科迪姆导数平方可积 :

$$\mathcal{L}_2(\pi) = \left\{ \mu \in \mathrm{SM}(\mathcal{X}) : \mu \ll \pi \text{ and } \int \left(\frac{d\mu}{d\pi} \right)^2 d\pi < \infty \right\},$$

其中，$\mathrm{SM}(\mathcal{X})$ 表示定义在 $\mathcal{X}$ 上的符号测度集合。测度 μ 对算子 P 的作用定义为

$$\mu P(A) = \int P(x, A)\, d\mu(x)$$

对所有可测集 A 均成立。这一性质在应用中至关重要，因为它量化了链收敛到平稳状态的速率。具体而言，几何遍历性保证，达到给定近似精度所需的步数随精度提高而按对数级增长。

马尔可夫链的转移核自然诱导出一个线性积分算子，称为马尔可夫算子；该算子通过对一步转移进行平均来作用于函数。如果一个算子的奇异值可求和，则称其为迹类，这一条件意味着该算子具有紧性。当链具有哈里斯遍历性时，紧性保证该算子的谱除位于 1 处的简单特征值外，均严格位于单位圆盘内，因此产生谱隙，进而保证几何遍历性 [22]。

针对高维线性模型的吉布斯采样器的几何遍历性，已通过李雅普诺夫函数以及次水平集上的小化条件得到确立 [30]。这一框架已应用于全局-局部收缩先验，包括贝叶斯套索[16, 23], 狄利克雷-拉普拉斯先验 [5, 21],和马蹄先验 [6, 11]。相关模型的迹类性质已在 [22, 35, 36] 中得到确立。

在本文中，我们从两个方面扩展了这些结果。首先，我们考虑以广义逆高斯（GIG）族作为混合分布，它包含此前研究过的许多先验分布，包括 [21] 的正态-伽马先验这一特殊情形。其次，我们采用比早期研究所考虑的标准线性回归设定更一般的模型结构，因此这些结果的适用范围更广。

在以下小节中，我们将介绍该模型类别，并以随机梯度下降（SGD）优化作为例子，说明为何推导该模型类别的几何遍历性很重要。

1.2. 潜在高斯模型及其非高斯扩展。层次模型是用于描述多个尺度间依赖关系的基础统计框架。潜在高斯模型（LGMs）是一大类层次模型，其中未观测到的潜变量服从高斯分布。这类模型涵盖广义线性混合模型 [14],、生存模型 [19],以及空间和时空模型 [3],等。

对线性高斯模型的推断可以通过从后验分布中进行马尔可夫链蒙特卡罗抽样来实现。对于涉及大规模潜在高斯场的问题，积分嵌套拉普拉斯近似（INLA）方法通常能提供更高效、准确的推断（[31]）。线性高斯模型具有如下层次结构:

$$\begin{aligned} \text{Data} \quad & Y|W, \theta_1, \theta_2 \sim \prod_{i \in \mathcal{I}} \pi(Y_i|W_i, \theta_1), \\ \text{Latent field} \quad & W|\theta_2 \sim N(\mu(\theta_2), Q^{-1}(\theta_2)), \\ \text{Hyperparameters} \quad & \theta_1 \sim \pi(\theta_1), \quad \theta_2 \sim \pi(\theta_2). \end{aligned} \tag{1}$$

该层次结构由三个阶段组成：(i) 观测模型，用于指定数据 Y 在给定潜变量 W 和超参数 θ_1条件下的条件分布；(ii) 在给定超参数 θ_2条件下的潜在高斯场 W；以及(iii)超参数先验分布。

尽管高斯过程广受欢迎且具有灵活性，但对于存在局部尖峰 [33], 和突变 [12]的数据集，它们可能过于平滑。线性高斯模型的非高斯扩展通过允许潜变量表现出偏度（非对称分布）、重尾（频繁出现极端值）、样本路径不连续性（突然跳跃）和方向不对称性（分布两侧的变异性不相等），从而克服高斯假设的局限性。这些特征可以提高预测准确度，并更好地刻画局部数据特征（[20], [7], [34]）。一种构建非高斯过程的方法在 [7] 中提出，并在 [2],中加以扩展，采用方差混合表示。

在本文中，我们具体考虑具有高斯似然的潜在高斯模型（LGMs）的以下非高斯扩展，并将其称为线性潜在非高斯模型（LLnGM）：

$$\text{Data} \quad Y|W,V \sim N(X\beta + AW, \sigma_\epsilon^2 I), \tag{2a}$$

$$\text{Latent fields} \quad K(\zeta)W|V \sim N(\mu V - \mu h, \text{diag}(\sigma^2 V)), \tag{2b}$$

$$\text{MixVars} \quad V_i \overset{\text{ind.}}{\sim} \text{GIG}(p,a,b). \tag{2c}$$

其中， $Y = (Y_1, \ldots\ldots, Y_m)^\top$是数据向量， $W = (W_1, \ldots\ldots, W_n)^\top$是潜在场，而$\zeta = (\zeta_1, \ldots\ldots, \zeta_p)^\top$是可逆方阵 K的超参数，该矩阵决定了具体的模型结构。此外， $\mu \in \mathbb{R}$ 和 $\sigma > 0$ 是控制潜在模型方差和偏度的参数， X 是设计矩阵，其对应的回归系数为 β。矩阵 A 将潜在场映射到观测空间，而 h是由具体模型选择决定的固定向量（下面将给出示例）。我们还可以通过为这些参数添加先验分布，使该框架具有贝叶斯特性，但这并非本文的主要关注点，也不会影响后续结果。

模型的最后一个组成部分是混合变量向量 $V = (V_1, \ldots, V_n)^\top$，假设这些变量相互独立且服从广义逆高斯分布（GIG）。GIG 密度函数为

$$f(x;p,a,b) = \frac{(a/b)^{p/2}}{2K_p(\sqrt{ab})} x^{p-1} \exp\left\{-\frac{1}{2}(ax + b/x)\right\}, \quad x > 0,$$

其中， K_p 是阶数为 p 的第二类修正贝塞尔函数，参数空间为 $\Psi = \Psi_I \cup \Psi_{IG} \cup \Psi_\Gamma$，其中

$$\Psi_I = \{a > 0, b > 0, p \in \mathbb{R}\}, \quad \Psi_{IG} = \{a = 0, b > 0, p < 0\}, \quad \Psi_\Gamma = \{a > 0, b = 0, p > 0\}.$$

这里， Ψ_I 是参数空间的内部， Ψ_{IG} 是对应于逆伽马分布的边界情形， Ψ_Γ 是对应于伽马分布的边界情形。广义逆高斯分布族还将逆高斯分布（$p = 1/2$）作为一种内部情形包含在内。一个重要性质是其缩放行为：若 $V \sim \text{GIG}(p,a,b)$，则对任意 $c > 0$, $cV \sim \text{GIG}(p, a/c, cb)$。

当广义逆高斯分布用作高斯方差混合中的混合分布时，所得边缘分布属于广义双曲（GH）族（[4]）。这是一类非常灵活的分布，其中包括双曲分布、Student's t 分布、广义非对称拉普拉斯分布（ GAL ）和正态逆高斯分布（ NIG ）等。具体而言，GH 随机变量 Z 可表示为

$$Z = \delta + \mu V + \sigma\sqrt{V}U, \tag{3}$$

其中， $\delta \in \mathbb{R}$、 $\mu \in \mathbb{R}$和 $\sigma > 0$ 分别为位置、漂移和尺度参数， $V \sim \text{GIG}(p,\ a,\ b)$，且 $U \sim N(0,\ 1)$与 V相互独立。完整的广义双曲分布有六个参数$(\delta,\ \mu,\ \sigma,\ p,\ a,\ b)$；然而，由于我们的模型包含能够确定位置的固定效应，为避免参数过度化，我们在式（2）的构造中固定 $\delta = -\mu h$ 。表1列出了通过改变 GIG 参数$(p,\ a,\ b)$得到的一些特殊情形。

表1： 广义双曲分布的特殊情形及其对应的广义逆高斯混合分布

X的分布	V的分布	V的广义逆高斯形式
t	逆伽马分布(ν， ν)	广义逆高斯分布($-\nu$， 0， 2ν)
NIG	逆高斯分布(α， β)	广义逆高斯分布(-0.5， α， β)
GAL	伽马分布(α， β)	广义逆高斯分布(α， 2β， 0)
柯西分布	逆伽马分布(0.5， ν)	广义逆高斯分布(-0.5, $0, 2\nu$)

通过为算子 K指定不同形式，该框架可以容纳多种广泛使用的潜变量非高斯模型，包括自回归模型、随机游走模型、SPDE Matérn 模型和时空模型；有关算子 K 及其对应模型的若干示例，请参见 [10]。对于这些模型中的每一个，向量 h 都有明确的表达式。例如，对于自回归模型， $h = 1$；而 h 则由 SPDE Matérn 模型 [7]的离散化确定。

非高斯潜在建模中反复面临的一项挑战是，边际行为越丰富，通常越要以失去可扩展的高斯计算工具为代价。LLnGMs 通过引入混合变量来规避这一权衡，使潜在场在给定条件下服从高斯分布。因此，仍可利用稀疏算子结构（例如自回归形式所产生的结构）进行高效的计算线性代数运算，而非高斯特征则局限于混合分量的一维更新中。

1.3. 随机梯度下降。一个自然的问题是如何利用数据拟合模型 (2)。在某些情形下，可以采用 [10] 的变分贝叶斯方法；此外，蒙特卡洛期望最大化（MCEM）方法也已用于该类中的某些非高斯模型 [7, 34]。不过，最广泛使用的方法是基于随机梯度下降（SGD）方法的最大似然估计， [参见例如 2, 8]。事实上，通用软件 ngme2（ https://github.com/davidbolin/ngme2）实现了许多形式为 (2) 的潜变量非高斯模型，并采用了这一方法。

随机梯度下降（SGD）由 [26], 提出，是一种广泛用于机器学习、神经网络和统计推断等大规模问题的迭代优化方法。使用随机梯度下降最小化函数 $f(\theta)$ 时，我们从初始参数向量 $\theta^{(0)}$ 开始，并通过以下方式迭代更新：

$$\theta^{(t+1)} = \theta^{(t)} - \gamma_t \widehat{\nabla f}(\theta^{(t)}), \tag{4}$$

其中， $\gamma_t > 0$ 是第 t 次迭代的步长， $\nabla f(\theta)$ 是满足 $\frown E[\nabla f(\theta)] = \nabla f(\theta)$ 的随机向量。在步长 γ_t 满足适当条件且假设相对宽松的情况下，迭代值几乎必然收敛到 f [9] 的一个极小值点。

在 LLnGMs 的背景下，函数 f 是负对数似然，相应随机向量 $Q_n(\theta)$ 通过吉布斯采样进行抽样。在这种情形下，吉布斯采样得以应用的关键性质是：给定 X 后， V 的条件分布仍服从广义逆高斯分布。详情见第 2.1 节和第 3 节。不过，由于估计过程中使用了吉布斯采样器，几何遍历性为使用小批量样本估计梯度提供了依据，也为估计量提供了理论保证。

需要注意的是，对于某些非高斯噪声模型，基于似然并通过随机梯度下降进行的估计可能会失效。尤其是，当潜在场由广义非对称拉普拉斯（GAL）噪声驱动时，相应的方差混合变量 V服从形状参数较小的伽马分布（即 $p < 1/2$）。如第3.3节所述，这会导致估计中出现严重的数值和统计困难。这一问题尤为

突出地出现在基于随机偏微分方程的模型中。随着网格细化，有效形状参数p会随离散化层级的提高而减小；对于连续模型中驱动噪声的任意固定 $p>0$，当网格不断细化时，离散化模型中噪声的 $p\to 0$ 成立 [34]。因此，混合分布会变得越来越退化，导致似然评估和梯度估计均不稳定。在 [34], 中也已观察到这一现象，并导致所提出的蒙特卡洛期望最大化算法出现稳定性问题。本文的一项关键贡献是，我们随后证明，通过将式（2）中的所谓中心化参数化改为下一节介绍的另一种非中心化参数化，可以避免这些问题。

1.4. 主要贡献与文章结构。 本文的主要贡献如下　s.

1. 我们证明了应用于 LLnGM 的吉布斯采样器具有几何遍历性和迹类性质，其中非高斯性通过广义逆高斯混合变量引入。结果适用于一般的（可逆）算子矩阵 $K(\zeta)$，并涵盖完整的 GIG 参数空间，包括 NIG、广义非对称拉普拉斯分布和 Student-t 特殊情形。

2. 我们通过 Fisher 恒等式分析随机梯度下降估计，并确定了可积性条件。这些条件在下文引入的非中心化参数化下清晰易懂，但在式（2）的原始中心化表述中则较为微妙。具体而言，在中心化参数化下，完整数据得分在后验分布下可能不可积，从而无法直接应用遍历定理，这意味着无法证明随机梯度下降方法收敛。我们表明，采用非中心化参数化可以避免这些问题，因此能够基于遍历性结果推导随机梯度下降方法的收敛性。

3. 我们给出数值实验，以展示所提出的吉布斯采样器的混合行为，并通过实证验证理论结果。

本文结构如下。第 2 节介绍吉布斯采样器并陈述主要结果。第3节介绍其在随机梯度下降中的应用。第 4 节给出主要结果的证明，第 5 节呈现验证这些结果的模拟研究。文章最后在第 6 节进行讨论，并在三个附录中提供技术细节。如前所述，R 包 ngme2 实现了适用于该类模型的所提估计方法，所有数值实验均可通过此软件包复现。

2. 吉布斯采样器与几何遍历性的主要结果。

2.1. 吉布斯采样器。在 LLnGM 中，我们有未观测到的混合变量 V 和潜在场 W。我们的目标是从未观测变量的后验分布中进行抽样，即 $V,\ W|Y$。由于没有闭式表达式，直接从全条件分布中抽样较为困难。不过，我们可以使用吉布斯采样器迭代地从条件分布中抽样。我们提出了算法 1 中给出的吉布斯采样算法。

式 (2) 中的表述（下文称为中心化参数化）不便于进行理论分析。为简化后续证明，我们通过引入变换后的潜在场，构造一种等价的非中心化参数化 。具体而言，我们定义辅助潜变量 $M:=K(\zeta)W+\mu h$，使得原始潜在场可通过双射映射 $W=K(\zeta)^{-1}(M-\mu h)$恢复。在此变换下，层次模型变为

(5a)　　Data　$Y\mid M,V\sim N\big(X\beta+AK(\zeta)^{-1}(M-\mu h),\ \sigma_\epsilon^2 I\big),$

(5b) Latent fields $M \mid V \sim N\big(\mu V,\ \mathrm{diag}(\sigma^2 V)\big)$,

(5c) Mixing variables $V_i \overset{\text{ind.}}{\sim} \mathrm{GIG}(p, a, b), \qquad i = 1, \ldots, n.$

由于对于每个固定的 ζ， W 与 M 之间的变换都是微分同胚，因此中心化和非中心化参数化会导出相同的边际似然 Y，从而定义等价的统计模型。此外，非中心化表示还通过改善后验分布下某些得分分量的可积性，简化了随机梯度推断（第3.3节）。对于式 (5) 中的非中心化参数化，吉布斯采样算法与之类似，只是采样的是 M，而非 W。

现在我们给出两种参数化所需的条件分布。为简化记号，我们固定一个超参数 ζ，并在记号中省略 ζ，也就是说，我们用 K 代替 $K(\zeta)$。

命题 2.1. 在中心化参数化 (2) 下， $W|V,Y$ 的条件分布为

$$W|V,Y \sim N\left(Q^{-1}\left(\frac{1}{\sigma^2}K^\top D_V^{-1}\mu(V-h) + \frac{1}{\sigma_\epsilon^2}A^\top(Y - X\beta)\right),\ Q^{-1}\right), \tag{6}$$

其中 $Q = \sigma^{-2}K^\top D_V^{-1}K + \sigma_\epsilon^{-2}A^\top A$ 和 $D_V = \mathit{diag}(V)$。

U在非中心化参数化 (5) 下， $M|V,Y$ 的条件分布为

$$M|V,Y \sim N\left(Q^{-1}\left(\sigma_\epsilon^{-2}B^\top(Y - X\beta + \mu Bh) + \frac{\mu}{\sigma^2}\mathbf{1}\right),\ Q^{-1}\right), \tag{7}$$

其中 $Q = \sigma_\epsilon^{-2}B^\top B + \sigma^{-2}D_V^{-1}$、 $B = AK^{-1}$ 和 $\mathbf{1}$ 是全 1 向量。

此外，在两种参数化下， V 的全条件分布都按分量分解，且仍属于广义逆高斯分布族。由于 $Y \mid W,V$ 和 $Y \mid M,V$ 不依赖于 V，因此有

$$V_i \mid W,Y \ \sim\ \mathrm{GIG}\left(p - \tfrac{1}{2},\ a + \frac{\mu^2}{\sigma^2},\ b + \frac{((KW)_i + \mu h_i)^2}{\sigma^2}\right), \qquad i = 1, \ldots, n, \tag{8}$$

and

$$V_i \mid M,Y \ \sim\ \mathrm{GIG}\left(p - \tfrac{1}{2},\ a + \frac{\mu^2}{\sigma^2},\ b + \frac{M_i^2}{\sigma^2}\right), \qquad i = 1, \ldots, n, \tag{9}$$

在 i 上相互独立。

证明见附录 B.1。

算法 1 LLnGM 的吉布斯采样器

1：对于中心化参数化， $U = W$。**对于非中心化参数化** $U = M$。**2：初始化** $V^{(0)}$。**3：对于** $t = 1$**，2，.....，** T**执行 4：从条件分布** $U|V^{(t-1)}Y$**中采样** $U^{(t)}$。**5：从条件分布** $V|U^{(t)}Y$**中采样** $V^{(t)}$。**6：结束循环**

2.2. 从线性算子的视角看吉布斯采样。LL-nGM 的吉布斯采样器诱导一条马尔可夫链 $(V^{(t)})$，其转移密度为:

$$k(V,\tilde{V}) = \int_{\mathbb{R}^n} \pi(\tilde{V}|W,Y)\pi(W|V,Y)dW, \tag{10}$$

其中 Y 为观测值，W 为潜在场，V 为混合场。令 $L_2(\pi)$ 表示关于目标后验密度 $\pi(V|Y)$ 平方可积的函数所构成的希尔伯特空间:

$$L_2(\pi) := \left\{ f : \mathbb{R}_+^n \to \mathbb{R}; \int_{\mathbb{R}_+^n} (f(V))^2 \pi(V|Y)dV < \infty \right\},$$

其中 $\mathbb{R}_+^n = \{V \in \mathbb{R}^n \colon V_i > 0, i = 1, \ldots, n\}$，配备标准内积:

$$(f,g) := (f,g)_{L_2(\pi)} = \int_{\mathbb{R}_+^n} f(V)g(V)\pi(V|Y)dV.$$

注意，$L_2(\pi)$ 与 $\mathcal{L}_2(\pi)$ 之间存在自然等同，其中每个函数$f \in L_2(\pi)$ 被等同于测度 $\mu_f(dV) = f(V)\pi(V|Y)dV = f(V)\pi(dV)$。马尔可夫转移密度 k（关于勒贝格测度）在 $L_2(\pi)$ 上诱导出一个线性算子Λ：

$$(\Lambda f)(V) = \int_{\mathbb{R}_+^n} k(V,\tilde{V}) f(\tilde{V}) d\tilde{V}. \tag{11}$$

现在我们导出算子 Λ 的基本性质。

命题 2.2（V-边缘马尔可夫算子的性质）。令 Λ 由(11) 定义，其转移密度 k 由(10)给出；令 π 表示具有密度 $\pi(\cdot \mid Y)$的 V-链的不变概率测度。则 Λ 满足 $\Lambda\mathbf{1} = \mathbf{1}$，并且是 $L_2(\pi)$上的压缩算子，即 $\|\Lambda\| \le 1$。此外，Λ 在 $L_2(\pi)$上是自伴且半正定的；因此$\mathrm{spec}(\Lambda) \subset [0,1]$。

证明。由于 $k(V,\tilde{V})$是马尔可夫转移密度，因此对于所有V，$k \ge 0$和$\int k(V,\tilde{V})\, d\tilde{V} = 1$，这意味着 $\Lambda\mathbf{1} = \mathbf{1}$。我们首先证明压缩性质。对于 π-几乎处处的 V，由詹森不等式可得，

$$(\Lambda f(V))^2 = \Big(\int k(V,\tilde{V}) f(\tilde{V})\, d\tilde{V} \Big)^2 \le \int k(V,\tilde{V}) f(\tilde{V})^2\, d\tilde{V}.$$

将两边对 $\pi(V \mid Y)\, dV$积分，并利用不变性$\int k(V,\tilde{V})\pi(V \mid Y)\, dV = \pi(\tilde{V} \mid Y)$，得到 $\|\Lambda f\|_{L_2(\pi)}^2 \le \|f\|_{L_2(\pi)}^2$。

接下来，我们证明自伴性和半正定性。对于 f，$g \in L_2(\pi)$，

$$\begin{aligned}
(\Lambda f, g)_{L_2(\pi)} &= \iint k(V,\tilde{V})\, f(\tilde{V})\, g(V)\, \pi(V \mid Y)\, d\tilde{V}\, dV \\
&= \iiint f(\tilde{V})\, g(V)\, \pi(\tilde{V} \mid W,Y)\, \pi(W \mid V,Y)\, \pi(V \mid Y)\, dW\, d\tilde{V}\, dV \\
&= \iiint f(\tilde{V})\, g(V)\, \pi(\tilde{V} \mid W,Y)\, \pi(V \mid W,Y)\, \pi(W \mid Y)\, dW\, d\tilde{V}\, dV \\
&= \int \pi(W \mid Y) \Big[\int f(\tilde{V}) \pi(\tilde{V} \mid W,Y)\, d\tilde{V} \Big] \Big[\int g(V) \pi(V \mid W,Y)\, dV \Big]\, dW \\
&= \mathbb{E}\Big[\mathbb{E}[f(V) \mid W,Y]\ \mathbb{E}[g(V) \mid W,Y] \ \Big|\ Y \Big],
\end{aligned}$$

它关于 $(f,\ g)$ 对称，从而证明了自伴性。令 $g=f$，得到

$$(\Lambda f,f)_{L_2(\pi)}=\mathbb{E}\left[\left(\mathbb{E}[f(V)\mid W,Y]\right)^2\ \middle|\ Y\right]\geq 0,$$

由此可证正半定性。最后，由于 Λ 是自伴的压缩算子，$\mathrm{spec}(\Lambda)\subset[-1,1]$，而正半定性进一步将 $\mathrm{spec}(\Lambda)\subset[0,1]$ 限定在更小范围内。 □

PROPOSITION 2.3（非周期性与哈里斯常返性）。假设转移密度˜ ˜满足 $k(V,V)>0$，对所有 $V,V\in\mathbb{R}^n_+$ 均成立。那么，V-边缘链是 ψ-不可约的、非周期的，且具有哈里斯常返性。

证明。 k 的严格正性意味着 ψ-不可约（例如，取 ψ 为 $\mathbb{R}^n_+$ 上的勒贝格测度），且链具有非周期性；参见例如 [29]。由于该链存在不变概率测度 $\pi(\cdot\mid Y)$，可知该链具有哈里斯常返性；参见 [32, 推论 2]。 □

根据 [27, 定理 2.1], ，当且仅当马尔可夫算子 Λ 在 ***1*** 处具有正谱隙时，Gibbs 链是几何遍历的：

$$\sup\bigl\{|\lambda|:\lambda\in\mathrm{spec}(\Lambda),\ \lambda\neq 1\bigr\}<1.$$

命题 2.2 表明，在 **(11)** 中定义的算子 Λ 是自伴的、半正定的，并且是 $L_2(\pi)$ 上的压缩算子，满足 $\Lambda\mathbf{1}=\mathbf{1}$；因此，$\mathrm{spec}(\Lambda)\subset[0,1]$，其中 1 是一个特征值。此外，命题 2.3 证明了该链是 Harris 常返的；结合唯一不变概率测度 $\pi(\cdot\mid Y)$的存在性，可知特征值1 是单重的。

命题 2.4（迹类算子具有 $L_2(\pi)$ 谱隙）。假设 Λ 是由 V-边缘转移密度 k 按照 (11) 在 $L_2(\pi)$ 上诱导的马尔可夫算子。如果 Λ 在 $L_2(\pi)$上属于迹类，则 Λ 在 $L_2(\pi)$ 的 1 处具有正谱隙；也就是说，

$$\lambda^\star:=1-\|\Lambda_0\|\in(0,1],\qquad \Lambda_0:=\Lambda\big|_{L_{2,0}(\pi)},\quad L_{2,0}(\pi):=\mathbf{1}^\perp.$$

因此，链$(V^{(t)})$ 在 $L_2(\pi)$中具有几何遍历性。此外，对于任意具有密度 $d\nu/d\pi\in L_2(\pi)$的概率测度 $\nu\ll\pi$，也就是说，对于任意概率测度$\nu\in\mathcal{L}_2(\pi)$，存在有限常数 $C(\nu)$ 使得

$$\|\nu\Lambda^m-\pi\|_{\mathrm{TV}}\leq C(\nu)\,(1-\lambda^\star)^m,\qquad m\geq 0.$$

证明。根据命题 2.2， Λ 是自伴的、半正定的，并且是$L_2(\pi)$上的压缩算子；因此 $\mathrm{spec}(\Lambda)\subset[0,1]$ 且 $\Lambda\mathbf{1}=\mathbf{1}$。由于 Λ 是迹类算子，因此必然是紧算子； consequently, 其谱仅由具有有限重数的特征值组成，0 是唯一可能的聚点。由于1 是一个单特征值（如上文所证），其余所有特征值均属于 $[0,1)$，从而得到 $\|\Lambda_0\|<1$，进而有 $\lambda^\star>0$。在 $L_2(\pi)$ 中的几何遍历性以及所述的全变差界均可由标准谱方法推出；例如，参见 [27,定理 2.1]。 □

我们的下一个目标是将迹类性与转移密度 k 的一个可积性条件联系起来；为此，我们引入标准化核

$$\hat{k}(V,\tilde{V}):=\frac{k(V,\tilde{V})}{\pi(\tilde{V}\mid Y)}.$$

˜这使我们能够将 Λ 的作用表示为关于测度

˜ ˜ π $=(dV)\pi(V|Y)dV$ 的积分:

$$(\Lambda f)(V)=\int f(\tilde{V})k(V,\tilde{V})d\tilde{V}=\int f(\tilde{V})\hat{k}(V,\tilde{V})\pi(d\tilde{V}).$$

根据 [25, 定理 2], Λ 在 $L_2(\pi)$ 上是迹类的，当且仅当

$$\int_{\mathbb{R}^n_+}\hat{k}(V,V)\pi(dV)<\infty,$$

这可归结为转移密度满足以下条件:

$$\int_{\mathbb{R}^n_+}k(V,V)\,dV<\infty. \tag{12}$$

根据命题 2.4，要确立 V-边际 Gibbs 链的 $L_2(\pi)$ 几何遍历性，只需验证可积性条件 (12)。为说明无论采用中心化表述还是非中心化表述，都可以进行这项验证，下面的命题将证明二者的转移核等价。

P命题 2.5（转移核的等价性）。中心化参数化和非中心化参数化所诱导的转移核相同；也就是说，对于所有 $V\in\mathbb{R}^n_+$ 和$Y\in\mathbb{R}^m$,

$$\int_{\mathbb{R}^n}\pi(\tilde{V}\mid W,Y)\pi(W\mid V,Y)\,dW=\int_{\mathbb{R}^n}\pi(\tilde{V}\mid M,Y)\pi(M\mid V,Y)\,dM.$$

证明。固定任意 $V\in\mathbb{R}^n_+$ 和 $Y\in\mathbb{R}^m$，并假设 $K(\zeta)$ 可逆。考虑双射

$$T:\mathbb{R}^n\to\mathbb{R}^n,\qquad M=T(W):=KW+\mu h,$$

并令 $|K|:=|\det K|$ 表示雅可比行列式的绝对值。在中心化的 ˜参数化中，给定(W,Y) 时，V的全条件分布满足

$$\pi(\tilde{V}\mid W,Y)\ \propto\ \pi(\tilde{V})\,\pi(W\mid\tilde{V})\,\pi(Y\mid W).$$

由于 $\pi(Y\mid W,V)$ 不依赖于 V，我们有 $\pi(Y\mid W,V)=\pi(Y\mid W)$；因此，˜ ˜$Y$和 V在给定 W的条件下相互独立，从而得到 $\pi(V\mid W,Y)=\pi(V\mid W)$。˜此外， $\pi(W\mid V)$对 W 的依赖仅通过涉及 $KW+\mu h$的潜在关系 ˜引入；因此， $\pi(V\mid W)$对 W的依赖仅通过 $M=KW+\mu h$。因此，由于 $\pi(V|W,V)$ 和 $\pi(V|M,V)$ 都是概率密度，我们有

$$\pi(\tilde{V}\mid W,Y)=\pi(\tilde{V}\mid M,Y)\qquad\text{whenever } M=KW+\mu h.$$

令 $\pi_W(\cdot\mid V,Y)$ 表示给定(V,Y)时 W的条件密度，并令 $\pi_M(\cdot\mid V,Y)$表示给定($V,Y)$时 $M=T(W)$ 的条件密度。根据变量替换公式,

$$\pi_M(M\mid V,Y)=\pi_W\big(T^{-1}(M)\mid V,Y\big)\,|K|^{-1},\qquad dM=|K|\,dW.$$

将此变换应用于核积分，我们得到

$$\begin{aligned}\int_{\mathbb{R}^n}\pi(\tilde{V}\mid W,Y)\,\pi_W(W\mid V,Y)\,dW&=\int_{\mathbb{R}^n}\pi(\tilde{V}\mid T(W),Y)\,\pi_W(W\mid V,Y)\,dW\\&=\int_{\mathbb{R}^n}\pi(\tilde{V}\mid M,Y)\,\pi_M(M\mid V,Y)\,dM\\&=\int_{\mathbb{R}^n}\pi(\tilde{V}\mid M,Y)\,\pi(M\mid V,Y)\,dM,\end{aligned}$$

由此得以证明所声称的等式。 □

2.3. 关于迹类性质与几何遍历性的主要结果。 本节介绍我们的主要结果，涉及 (i) 式 (11) 中定义的 V 边缘马尔可夫算子 Λ 的迹类性质，以及 (ii) 由吉布斯采样器诱导的马尔可夫链 $(V^{(t)})$ 的几何遍历性。完整证明推迟至第 4 节。

定理 2.1（迹类性质）。考虑采用中心化或非中心化参数化的 *LLnGM* 吉布斯采样器。式(11) 中定义的马尔可夫算子 Λ 在 $L_2(\pi)$ 上是迹类的，只要满足以下任一条件：

1. $a>0$、 $b>0$ 和 $p\in\mathbb{R}$;
2. $a>0$、 $b=0$ 和 $p>\frac{1}{2}$。

对于定理 2.1 未涵盖的参数区域，我们将通过漂移-小集化论证证明几何遍历性。为陈述这一结果，我们先引入以下假设。

假设 2.1（零空间小性假设）。考虑第 1.2 节中引入的中心化或非中心化参数化，且 K 可逆。令 $r:=\dim\mathrm{Null}(A)$。若 $r>0$，令 $U_A\in\mathbb{R}^{n\times r}$ 为 $\mathrm{Null}(A)$ 的标准正交基，并定义

$$G:=U_A^\top K^\top K\,U_A\in\mathbb{R}^{r\times r},\qquad z:=U_A^\top K^\top\mathbf{1}\in\mathbb{R}^r.$$

我们假设 $r=0$ 成立，或者

$$\frac{|\mu|}{\sqrt{\sigma^2 a+\mu^2}}\,\sqrt{z^\top G^{-1}z}<1. \tag{13}$$

请注意，当 $\mu=0$ 时，此条件会自动满足。

备注。式 (13) 中的零空间小性比率对漂移方向 1、先验结构矩阵 K 与设计矩阵 A 的零空间之间的相互作用施加了几何约束。术语 $\sqrt{z^\top G^{-1}z}$ 量化了变换后的向量 $K^\top\mathbf{1}$ 在由受限 Gram 矩阵 $G=U_A^\top K^\top K\,U_A$ 在 零空间(A) 上诱导的几何结构下，向子空间 零空间(A) 投影的大小。具体而言，$z=U_A^\top(K^\top\mathbf{1})$ 表示 $K^\top\mathbf{1}$ 投影到 零空间(A) 的基向量上所得的系数，而 $G=U_A^\top K^\top KU_A$表示由 K 变换后的基向量的 Gram 矩阵，刻画了先验精度矩阵限制在零空间上的曲率。因此，对维数n 的任何依赖，都取决于向量 $K^\top\mathbf{1}$ 与 零空间(A) 的对齐程度相对于 $K^\top K$ 在该子空间上的条件数而言有多强。

为保证可扩展性，具体而言，要使项 $\sqrt{z^\top G^{-1}z}$ 在 $n\to\infty$ 时保持有界，只需漂移方向不会越来越集中于不可识别子空间。例如，以下结构性假设足以保证与维数无关的行为：

(i) 常数模态的正交性： $\mathbf{1}\perp$ 零空间(A) （等价地， $\mathbf{1}\in$ 值域$(A^\top)$），这意味着 $U_A^\top\mathbf{1}=0$。(ii) 保持 的结构： 矩阵 K 不会将常数向量完全映射到零空间中。具体而言，假设 $K^\top\mathbf{1}=c\,\mathbf{1}+r$，其中 c 是一个标量，余项 r 在零空间上的投影有界：$\|U_A^\top r\|_2=O(1)$。(iii) 一致非退化性：将矩阵 K 限制在零空间(A)上时，它仍然条件良好，即 $\lambda_{\min}(G)=\lambda_{\min}(U_A^\top K^\top KU_A)\geq c_0>0$ 关于 n一致成立。

表2 吉布斯采样器在不同参数范围下的几何遍历性与迹类性质

参数范围	迹类	几何遍历性
$a>0, b>0, p\in\mathbb{R}$	Yes	Yes
$a>0, b=0, p>1/2$	Yes	Yes
$a>0, b=0, 0<p\le 1/2$	未知	是* (需要假设 2.1)
$a=0, b>0, p<0, \mu=0$	No	Yes
$a=0, b>0, p<0, \mu\neq 0$	未知	是* （需要满足假设 2.1）

在确立零小性条件后，我们现在陈述漂移-小集框架下边缘 V-链几何遍历性的主要结果。证明见第 4.2 节。

命题 2.6（通过漂移-小集条件证明几何遍历性）。考虑由 (25) 在 $V\in(0,\infty)^n$上导出的吉布斯采样器。对于所考虑的高斯模型，条件密度 $\pi(M\mid V,Y)$ 处处为正，且关于 V 连续，因此以下结论成立。回顾一下， $\tilde{a}:=a+\mu^2/\sigma^2$。在以下各参数区域中，边缘链 $(V_{(t)})_{t\ge 0}$ 均具有几何遍历性：

~~*(I)* $a=\mu=0$ （等价地， $\tilde{a}=0$）。在此状态下，必有 $p<0$ 和 $b>0$，且 $V_i\mid M_i$服从逆伽马分布。该结论无条件成立。*(II)* $a=0, b>0, p<0$和 $\mu\neq 0$ （因此 $\tilde{a}=\mu^2/\sigma^2>0$），并满足假设 *2.1*。对于~~ ~~$a=0$~~，假设 *2.1* 化简为 $\sqrt{z^\top G^{-1}z}<1$。在此状态下， $\tilde{V}_i\mid M_i$ 服从参数为 $(p-\frac{1}{2}, \tilde{a}, b+(M_i/\sigma)^2)$ 的广义逆高斯分布。*(III)* $a>0, b=0, p>0$（因此 $\tilde{a}=a+\mu^2/\sigma^2>0$），并满足假设 *2.1*。在此状态下，$\tilde{V}_i\mid M_i$服从参数为 $(p-\frac{1}{2}, \tilde{a}, (M_i/\sigma)^2)$ 的广义逆高斯分布。

前述结果——定理2.1和命题 2.6——共同刻画了参数空间互补区域上的几何遍历性。现在，我们将这些发现综合为一项完整的陈述，从而全面展示所有参数区域中的几何遍历性（表 2）。

定理 2.2（全参数空间上的几何遍历性）。考虑由 (25) 在 $V\in(0,\infty)^n$ 上诱导的吉布斯采样器，并假设 $\pi(M\mid V,Y)$ 在 V 上处处为正且连续。则边缘链 $(V_{(t)})_{t\ge 0}$ 在表 *2* 所列的每个参数区域中均具有几何遍历性。

证明 . 证明按照表中各状态进行分类讨论
2。对于情形 $a>0, b>0, p\in\mathbb{R}$和 $a>0, b=0, p>\frac{1}{2}$，定理 2.1 证明算子 Λ 是迹类，因此由命题 2.4 可得几何遍历性。

对于其余状态，我们采用命题 2.6 中的漂移-小集条件结果。具体而言：情形 (I) 涵盖 $a=0$、$b>0$、$p<0$、$\mu=0$；情形 (II) 在假设 2.1 下涵盖 $a=0$、$b>0$、$p<0$、$\mu\neq 0$；情形 (III) 在假设 2.1 下涵盖 $a>0$、$b=0$、$p\le\frac{1}{2}$。 □

一个重要的特殊情形是 NIG模型，我们将其作为一个推论陈述。

推论 2.1（NIG 场的几何遍历性）。假设混合变量服从逆高斯分布， $V_i\sim IG(a,b)$，其中 $a,b>0$对应于

$p=-1/2$ *GIG* 参数化中的参数。则潜在场 W 服从 *NIG* 分布，算子 Λ 是迹类，并且对于漂移参数 μ的所有取值，吉布斯采样器在 $L_2(\pi)$ 意义下具有几何遍历性。

证明。由于 $V_i \sim$ IG（a，b）意味着 $a>0$、$b>0$ 和 $p=-1/2$，因此该结果直接由定理2.1（情形 1）和命题 2.4 得出。 □

3. 随机梯度估计。本节说明如何使用第 2.1 节中构造的吉布斯采样器，为 LLnGMs 的最大似然推断构造随机梯度估计量。

3.1. 通过费舍尔恒等式计算似然梯度。令 $\theta \in \Theta$ 表示所有模型参数组成的向量，并令 p_θ（Y，Z）表示观测数据 Y 与潜变量 Z 的联合密度。如第 2.1 节所述，我们可以采用带有 $Z=$（W，V）的中心化参数化，或采用带有 $Z=$（M，V）的非中心化参数化。观测数据的边际似然为

$$p_\theta(Y)=\int p_\theta(Y,Z)\,\lambda(dZ),$$

其中，λ 是潜在空间上的支配测度。相应的对数似然为ℓ（θ；Y）：$=$对数p_θ（Y）。

直接最大化 $\ell(\theta;Y)$ 颇具挑战，因为定义 $p_\theta(Y)$ 的积分通常难以计算，因此无法得到对数似然的闭式表达式。为使用基于梯度的优化进行最大似然估计，我们转而寻求计算梯度 $\nabla_\theta \ell(\theta;Y)$。尽管对数似然本身难以计算，其梯度可通过费舍尔恒等式表示为后验期望值。

命题 3.1（费舍尔恒等式 [13, 附录 D。2]）。假设 $p_\theta(Y,Z)$ 关于 θ 可微，且微分与积分的顺序可以交换（满足标准的控制收敛条件）。那么，

$$\nabla_\theta \ell(\theta;Y)=\int \nabla_\theta \log p_\theta(Y,Z)\,p_\theta(Z\mid Y)\,\lambda(dZ)=\mathbb{E}_\theta\big[\nabla_\theta \log p_\theta(Y,Z)\mid Y\big].$$

由于观测模型和潜在场具有条件高斯结构，且 V服从参数混合分布，完整数据对数似然对数 $p_\theta(Y,Z)$ 可以写成闭式形式。然而，命题 3.1 中的期望值在解析上仍难以计算。因此，我们使用从后验分布中抽取的样本，通过蒙特卡罗积分对其进行近似。

具体而言，设 $\{(V_{(j)}W_{(j)})\}_{j=1}^{k}$为由第 2.1 节所述的吉布斯采样器生成的联合后验分布 $p_\theta(V,W\mid Y)$的样本。我们将完整数据得分函数定义为

$$g_\theta(Y,V,W):=\nabla_\theta \log p_\theta(Y,V,W).$$

$\nabla_\theta \ell(\theta;Y)$ 的自然蒙特卡罗估计量为

$$\widehat{\nabla \ell}_k(\theta;Y)=\frac{1}{k}\sum_{j=1}^{k} g_\theta(Y,V^{(j)},W^{(j)}). \tag{14}$$

在非中心化参数化下，使用吉布斯样本$(V_{(j)}M^{(j)})$以及相应的完整数据得分 ∇_θ对数 $p_\theta($ $Y,V,M)$，即可构造类似的估计量。

3.2. Rao–Blackwell化梯度估计量与马尔可夫链遍历性。第 4.2 节推导的几何遍历性适用于吉布斯采样器的 V-边际分布。需要强调的是，在每次应用吉布斯采样器的过程中，参数向量 θ 都保持固定。也就是说，对于随机梯度过程中的给定迭代 $\theta^{(t)}$，吉布斯采样器的目标是后验分布 $p_{\theta^{(t)}}(V, W \mid Y)$（或非中心化参数化下的 $p_{\theta^{(t)}}(V, M \mid Y)$）。参数 θ 仅在吉布斯采样阶段之间通过随机梯度更新进行更新，而不会在马尔可夫链本身的运行过程中更新。因此，所有遍历性和收敛结果都适用于给定 θ 固定值的条件情形，这正是下文所采用的马尔可夫链理论所要求的。这一理论保证确保链 $\{V^{(j)}\}$ 以几何速率收敛至目标边际后验分布 $p_\theta(V \mid Y)$。然而，式 (14) 中的基本蒙特卡罗估计量 $\widehat{\nabla \ell}_k(\theta;\, Y)$ 涉及联合样本 $\{(V^{(j)}, W^{(j)})\}$。为了严格应用针对 V-链所建立的遍历极限定理，理想的做法是构造一个仅依赖序列 $\{V^{(j)}\}$ 的梯度估计量。幸运的是，条件模型结构使我们能够在给定 (V, Y) 的情况下，对潜在场 W（或 M）进行解析积分，从而得到一个仅依赖 V 的函数。这个过程称为 Rao–Blackwell化，不仅使估计量与我们关于 V 的理论结果相吻合，还能降低估计量的方差。

根据全方差公式，对子σ代数进行条件化会降低估计量的方差：若 $\delta(X)$ 是可积随机变量，则

$$\operatorname{Var}\big(\mathbb{E}[\delta(X) \mid Y]\big) \le \operatorname{Var}(\delta(X)).$$

在我们的设定中，给定 (V, Y)后，潜在场 W （或非中心化参数化中的 M ）服从高斯分布。借助这一结构，我们可以通过解析积分消去高斯层，从而定义 Rao–Blackwell 化得分：

$$g_\theta^{\mathrm{RB}}(Y, V) := \mathbb{E}_\theta[\nabla_\theta \log p_\theta(Y, V, W) \mid Y, V]\,.$$

根据塔式性质，该得分仍是梯度的无偏估计量：

$$\mathbb{E}_\theta\big[g_\theta^{\mathrm{RB}}(Y, V) \mid Y\big] = \mathbb{E}_\theta[\nabla_\theta \log p_\theta(Y, V, W) \mid Y] = \nabla_\theta \ell(\theta; Y).$$

因此，我们只需使用吉布斯样本中 V 分量 $\{V^{(j)}\}_{j=1}^k$，即可构造 Rao–Blackwell 化梯度估计量：

$$\widehat{\nabla \ell}_k^{\mathrm{RB}}(\theta; Y) = \frac{1}{k} \sum_{j=1}^{k} g_\theta^{\mathrm{RB}}(Y, V^{(j)}). \tag{15}$$

一个关键的计算优势是，g_θ^{RB}（Y，V）通常可以表示为仅涉及 $W \mid$（V，Y）的条件均值和协方差的闭式表达式，从而无需对 W 进行采样来进行梯度估计（但在对 V进行吉布斯采样时，W 仍是其中一个步骤）。

该估计量的收敛性由马尔可夫链的遍历性质保证。假设链 $\{V^{(j)}\}$ 是（几何）遍历的，具有不变分布 p_θ（$V \mid Y$），且得分函数满足可积性条件

$$\mathbb{E}_\theta\big[\big\|g_\theta^{\mathrm{RB}}(Y, V)\big\| \mid Y\big] < \infty. \tag{16}$$

那么，马尔可夫链遍历定理确保几乎必然收敛：

$$\widehat{\nabla \ell}_k^{\mathrm{RB}}(\theta; Y) \longrightarrow \mathbb{E}_\theta\big[g_\theta^{\mathrm{RB}}(Y, V) \mid Y\big] = \nabla_\theta \ell(\theta; Y), \qquad \text{as } k \to \infty.$$

一个微妙但重要的点是，在实践中，梯度估计量$g_\theta^{\mathrm{RB}}(Y, V)$ 是使用针对 $p_\theta(V \mid Y)$的有限长度吉布斯采样器来计算的。除非马尔可夫链严格在平稳性下初始化，否则对于有限次吉布斯运行，所得梯度估计并非精确无偏，而是渐近无偏。残余偏差来自该链的有限步边缘分布与目标平稳分布之间的偏离。

尽管如此，所得随机梯度方案的收敛仍可得到保证。经典随机逼近理论允许有偏的梯度估计量，前提是偏差相对于步长序列衰减得足够快。具体而言，若随机梯度下降更新允许如下分解

$$\theta^{(t+1)} = \theta^{(t)} - \gamma_t\big(\nabla f(\theta^{(t)}) + \xi_t + b_t\big),$$

其中 ξ_t 是鞅差噪声项， b_t 是由马尔可夫链蒙特卡罗收敛不充分所导致的偏差；只要偏差序列满足可和条件 $\sum_{t=1}^{\infty}\gamma_t\|b_t\| < \infty$ [17, , 就能保持几乎必然收敛（定理 2）。1]。

在当前情境下，我们依赖吉布斯采样器的几何遍历性。在温和的正则性假设下，例如满足矩条件以充分控制蒙特卡洛误差，通过漂移条件建立的几何遍历性已知可使有限长度马尔可夫链蒙特卡罗运行所导致的偏差呈几何衰减；此类代表性条件的示例见 [28, 事实 10] 。具体而言，经过 k 次吉布斯迭代后，梯度估计量的偏差满足

$$\|b_t\| \le M(\theta^{(t)}, V_{\text{init}})\,\rho^k,$$

其中 $\rho < 1$ 表示压缩率， M 是依赖于当前参数值和马尔可夫链初始化的有限常数。因此，在满足 $\sum_t \gamma_t = \infty$ 和 $\sum_t \gamma_t^2 < \infty$ 的标准步长方案下，马尔可夫链蒙特卡罗偏差的累积影响渐近可忽略，不会影响随机梯度算法的极限动力学。

3.3. 得分函数的可积性与参数化的作用。 为确保梯度估计量几乎必然收敛，不变分布 $p_\theta(V \mid Y)$必须满足可积性条件 **(16)**。尽管无论采用何种潜在场参数化， V 的边缘链都相同，但中心化表示与非中心化表示下得分 $g_\theta^{\mathrm{RB}}(Y, V)$ 的函数形式却有显著差异。这一选择很重要，因为如果得分函数存在奇点，或在 V 的定义域中存在实际上未定义的区域，那么验证可积性可能颇具挑战。

在中心化 参数化下，完整数据得分向量本身就包含与混合变量的逆幂成比例的项。具体而言，关于 μ 的得分为

$$\partial_\mu \log p_\theta(Y, V, W) = \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[(KW)_i - \mu V_i + 2\mu h_i - h_i\frac{(KW)_i + \mu h_i}{V_i}\right],$$

因此，验证正则性条件 **(16)** 需要控制混合变量负阶矩的后验期望（两种参数化下所有得分分量的显式表达式见附录 C）。下面的引理给出了这些后验期望存在的一个精确条件，并揭示了当$\alpha \le 1/2$时出现的可积性问题。

引理 3.1. 固定 $i \in \{1, \dots, n\}$，并假设 K 可逆。在条件$0 < \pi(y) < \infty$下，后验期望值

$$\mathbb{E}\left[\left|\frac{(KW)_i + \mu h_i}{V_i}\right| \,\middle|\, Y = y\right]$$

*is*有限当且仅当 $\alpha > 1/2$。特别是，只要 $\alpha \le 1/2$，该期望就会发散 .

证明见附录 B.2。该引理表明，这种障碍并非仅仅源于先验：只要似然在 $V_i \approx 0$ 附近没有充分抑制概率质量，且在系数 $|h_i\big(\ （\ KW\ ）\ _i + \mu h_i\big)|$ 远离零的事件上，中心化得分分量的后验期望就会变为无穷大。在中心化表述中，这两个条件在许多情形下都自然成立，因为当 $V_i \to 0$ 时，似然未必会迅速衰减，而系数 $h_i\big(\ （KW）\ _i + \mu h_i\big)$ 通常在具有正概率的集合上非退化。因此，正则性条件 (16) 并非自动成立；若没有额外的模型特定假设来约束似然在原点附近的行为，该条件可能不成立。

然而，第 2.1 节引入的非中心化 参数化可以用来规避这些困难。这种表示的一个关键优势在于，相对于 V，由此得到的得分函数通常具有更简单的代数结构。特别是，梯度不涉及 V 的奇异负次幂；相反，它们通常呈多项式形式。例如，非中心化参数化下对应的完整数据得分为

$$\partial_\mu \log p_\theta(Y, V, M) = \frac{1}{\sigma^2} \sum_{i=1}^{n} (M_i - \mu V_i) + \frac{1}{\sigma_\epsilon^2} (AK^{-1}h)^\top \left(Y - X\beta - AK^{-1}(M - \mu h)\right).$$

由于标准漂移条件（例如引理 4.11 和引理 4.15 中的条件）可以直接给出 V 多项式矩的界，因此验证可积性条件变得十分简单。基于此，我们采用非中心化参数化，从理论上证明随机梯度估计量的收敛性。

我们指出，引理 3.1 针对的是标准梯度估计量。由于 Rao–Blackwell 化估计量本身具有较低的方差，预计它在两种参数化下都表现更好。虽然这种方差缩减有时可以缓解可积性问题，但我们更倾向于改变参数化的策略，因为这种方法能从结构上解决问题。

3.4. 总结。 最后，我们总结第 2.1 节中提出的吉布斯采样器如何为 LLnGM 的最大似然推断提供有理论依据的随机梯度下降估计。

费舍尔恒等式（第 3.1 节）通过将对数似然梯度表示为完整数据得分的条件期望，为基于似然的推断与后验模拟建立了关键联系。尽管边际似然在解析上难以处理，但可以利用潜变量的后验样本估计其梯度，从而为基于梯度的优化和基于马尔可夫链蒙特卡罗的最大似然估计奠定基础。

利用层次高斯结构，还可以通过在给定混合变量的条件下对潜在高斯场进行解析积分，构造 Rao-Blackwell 化梯度估计量（第 3.2 节）。所得估计量仅依赖于V，降低了蒙特卡罗方差，并与针对 V 边际 Gibbs 链所建立的几何遍历性结果自然契合。在温和的可积性条件下，遍历定理保证梯度估计几乎处处收敛。

关键的是，第3.3节表明，这种收敛是否成立取决于得分函数的可积性，而得分函数又会受到所选参数化方式的显著影响。尽管 V 的边缘分布保持不变，中心化参数化通常会使得分函数包含 V 的逆幂，因此难以验证其可积性，在某些情况下甚至不满足可积性。相比之下，非中心化参数化使得得分函数对 V 呈多项式依赖，因此可积性可由标准漂移条件直接推出。

总而言之，非中心化参数化不仅是一种计算工具，更是一个关键的理论要素，使得能够对 LLnGM 的可扩展最大似然推断进行严格论证的 Rao–Blackwell 化随机梯度下降估计。

4. 主要结果的证明。下文中，对于对称矩阵 A 和 B，我们用$B \succeq A$（分别用 $B \succ A$）表示洛温纳（分别为严格洛温纳）偏序，即$B-A$ 为半正定（分别为正定）。

4.1. 迹类性质证明。我们首先陈述证明定理2.1所需的三个技术引理；其证明推迟至附录B。

引理4.1. 设 A， $B \in \mathbb{R}^{n\times n}$ 为对称正定矩阵，且 u， $v \in \mathbb{R}^n$。那么

$$u^\top A u + v^\top B v - (u+v)^\top (A^{-1}+B^{-1})^{-1}(u+v) = (Au - Bv)^\top (B+A)^{-1}(Au - Bv).$$

特别地，$u^\top A u + v^\top B v \geq (u+v)^\top (A^{-1}+B^{-1})^{-1}(u+v)$。

引理4.2. 对于任意 $p \in \mathbb{R}$，存在一个常数 $C>0$，使得对于所有 $x>0$,$K_p(x) > C x^{-|p|}(1+x)^{-1/2+|p|}e^{-x}$。

引理 4.3. 设 $p \geq 0$ 与 $\gamma \geq 0$。对任意固定向量 $b \in \mathbb{R}^n$，以及对所有对称正定矩阵 A，存在常数 $C = C(p,n,\|b\|) > 0$，与 A 相互独立，使得

$$\int \Big(\prod_{i=1}^n (\gamma + x_i^2)^p\Big) \exp\Big\{-\tfrac{1}{2}(x^\top A x - 2b^\top x)\Big\}\, dx \\ \leq C\,|A|^{-1/2} \exp\Big\{\frac{1}{2} b^\top A^{-1} b\Big\} \prod_{i=1}^n \Big(\gamma^p + [A^{-1}]_{ii}^p \big(1 + \sum_{j=1}^n [A^{-1}]_{jj}^p\big)\Big).$$

我们现在着手证明定理2.1。

定理2.1的证明。在整个证明过程中，C 表示一个一般正数常数，与 V 和 M 相互独立，其取值可能逐行改变。不失一般性，我们假设 $\sigma = \sigma_\epsilon = 1$。设 $\tilde{p} = p - \frac{1}{2}$ 与 $\tilde{a} = a + \mu^2 > 0$。回顾 $B = AK^{-1}$，我们定义如下量： $D_V = \mathrm{diag}(V_1,\ldots,V_n)$、 $Q = B^\top B + D_V^{-1}$ 和 $m = B^\top x + \mu \mathbf{1}$，其中 $x := Y - X\beta + \mu B h$ 是一个固定向量。回顾 M 的条件分布为

$$\pi(M \mid V, Y) = C\,|Q|^{1/2} \exp\Big\{-\tfrac{1}{2}(M - Q^{-1}m)^\top Q (M - Q^{-1}m)\Big\},$$

并且，在给定 M 的条件下， V_i 的各分量相互独立，其密度为

$$\pi(V \mid M, Y) = C \prod_{i=1}^n \left[\frac{\big(\sqrt{\tilde{a}(b+M_i^2)}\big)^{-\tilde{p}}}{K_{\tilde{p}}\big(\sqrt{\tilde{a}(b+M_i^2)}\big)} V_i^{\tilde{p}-1}\right] \exp\left\{-\tfrac{1}{2}\Big(\tilde{a}\,\mathbf{1}^\top D_V \mathbf{1} + \sum_{i=1}^n \frac{b+M_i^2}{V_i}\Big)\right\}.$$

情形 ***1***： $a>0$， $b>0$， $p \in \mathbb{R}$。 根据引理 4.2，对于固定的 $\tilde{p} \in \mathbb{R}$，存在一个常数 $C>0$，使得对于所有 $t>0$、 $K_{\tilde{p}}(t) \geq C t^{-|\tilde{p}|}(1+t)^{-1/2+|\tilde{p}|}e^{-t}$。令 $t_i = \sqrt{\tilde{a}(b+M_i^2)}$，则有

$$\frac{t_i^{-\tilde{p}}}{K_{\tilde{p}}(t_i)} \leq C t_i^{|\tilde{p}|-\tilde{p}} (1+t_i)^{1/2-|\tilde{p}|} e^{t_i}.$$

定义指数 $\eta := (|\tilde{p}| - \tilde{p})/2 \geq 0$ 和 $\bar{\eta} := \max(1/4 - |\tilde{p}|/2, 0) \geq 0$。利用界$1 + t_i \leq C(1 + b + M_i^2)^{1/2}$和 $t_i \leq \sqrt{\tilde{a}}b + \sqrt{\tilde{a}}|M_i|$，并注意到 $t_i^{|\tilde{p}|-\tilde{p}} = (\tilde{a}(b+M_i^2))^{\eta}$，可得

$$\frac{(\sqrt{\tilde{a}(b+M_i^2)})^{-\tilde{p}}}{K_{\tilde{p}}(\sqrt{\tilde{a}(b+M_i^2)})} \leq C(b+M_i^2)^{\eta}(1+b+M_i^2)^{\bar{\eta}} \exp\{\sqrt{\tilde{a}}|M_i|\}.$$

将此结果代入 $\pi(V \mid M, Y)$ 的表达式，得到

$$\begin{aligned}\pi(V \mid M, Y) \le C\Big(\prod_{i=1}^{n} V_i^{\tilde{p}-1}(b+M_i^2)^{\eta}(1+b+M_i^2)^{\bar{\eta}}\Big) \\ \times \exp\Big\{-\tfrac{1}{2}\Big(\tilde{a}\,\mathbf{1}^{\top}D_V\mathbf{1}+b\,\mathbf{1}^{\top}D_V^{-1}\mathbf{1}+M^{\top}D_V^{-1}M\Big)+\sqrt{\tilde{a}}\sum_{i=1}^{n}|M_i|\Big\}.\end{aligned} \tag{17}$$

令 $\Phi := \{\pm 1\}^n$，注意到 $\sum_i |M_i| =$ 最大值$_{\phi\in\Phi}\,\phi^{\top}M$。利用不等式$\exp(\max_j u_j) \le \sum_j \exp(u_j)$，可得

$$\exp\Big(\sqrt{\tilde{a}}\sum_i |M_i|\Big) \le \sum_{\phi\in\Phi}\exp\{\sqrt{\tilde{a}}\,\phi^{\top}M\}.$$

将此界与 (17) 以及 $\pi(M \mid V, Y)$ 的表达式结合起来，我们确定了每个 $\phi\in\Phi$ 对应的 M 相关指数为 $-\frac{1}{2}M^{\top}(Q+D_V^{-1})M+(m+\sqrt{\tilde{a}}\phi)^{\top}M$。由此得到联合密度的以下界：

$$\begin{aligned}\pi(V \mid M, Y)\pi(M \mid V, Y) \le C\,|Q|^{1/2}\Big(\prod_i V_i^{\tilde{p}-1}(b+M_i^2)^{\eta}(1+b+M_i^2)^{\bar{\eta}}\Big) \\ \times \exp\Big\{-\tfrac{1}{2}\Big(\tilde{a}\,\mathbf{1}^{\top}D_V\mathbf{1}+b\,\mathbf{1}^{\top}D_V^{-1}\mathbf{1}+m^{\top}Q^{-1}m\Big)\Big\} \\ \times \sum_{\phi\in\Phi}\exp\Big\{-\tfrac{1}{2}\Big(M^{\top}\tilde{Q}\,M-2(m+\sqrt{\tilde{a}}\phi)^{\top}M\Big)\Big\},\end{aligned} \tag{18}$$

其中 $\tilde{Q} := Q+D_V^{-1} = B^{\top}B+2D_V^{-1} \succ 0$。

固定 $\phi\in\Phi$，并令 $d_\phi := m+\sqrt{\tilde{a}}\phi$。定义 $f(M) := \prod_{i=1}^{n}(b+M_i^2)^{\eta}$ 和 $g(M) := \prod_{i=1}^{n}(1+b+M_i^2)^{\bar{\eta}}$，应用柯西–施瓦茨不等式可得：

$$\begin{aligned}\int f(M)g(M)e^{-\frac{1}{2}(M^{\top}\tilde{Q}M-2d_\phi^{\top}M)}dM \le \Big(\int f(M)^2 e^{-\frac{1}{2}(M^{\top}\tilde{Q}M-2d_\phi^{\top}M)}dM\Big)^{1/2} \\ \times\Big(\int g(M)^2 e^{-\frac{1}{2}(M^{\top}\tilde{Q}M-2d_\phi^{\top}M)}dM\Big)^{1/2}.\end{aligned}$$

现在，我们将引理 4.3（取 $\gamma=b, p=2\eta$ 和 $\gamma=1+b, p=2\bar{\eta}$）应用于高斯核 $\exp\{-\frac{1}{2}(M^{\top}\tilde{Q}M-2d_\phi^{\top}M)\}$ 下的每个因子。由于 $Q\succeq D_V^{-1}$ 蕴含 $Q^{-1}\preceq D_V$，且$\tilde{Q}\succeq 2D_V^{-1}$ 蕴含 $\tilde{Q}^{-1}\preceq\frac{1}{2}D_V$，因此对所有 i，都有 $[\tilde{Q}^{-1}]_{ii}\le V_i$。由此可得：

$$\begin{aligned}\int\prod_i(b+M_i^2)^{\eta}(1+b+M_i^2)^{\bar{\eta}}\exp\Big\{-\tfrac{1}{2}(M^{\top}\tilde{Q}M-2d_\phi^{\top}M)\Big\}\,dM \\ \le C\,|\tilde{Q}|^{-1/2}\exp\Big\{\tfrac{1}{2}d_\phi^{\top}\tilde{Q}^{-1}d_\phi\Big\}\times S_\eta^{n/2}S_{\bar{\eta}}^{n/2}\prod_i(b^{\eta}+V_i^{\eta})\big((1+b)^{\bar{\eta}}+V_i^{\bar{\eta}}\big),\end{aligned} \tag{19}$$

其中 $S_\eta := 1+\sum_{k=1}^{n}V_k^{2\eta}$ 且 $S_{\bar{\eta}} := 1+\sum_{k=1}^{n}V_k^{2\bar{\eta}}$。对 M 积分，对 $\phi\in\Phi$ 求和，并利用行列式不等式

$$|Q|^{1/2}|\tilde{Q}|^{-1/2} = \big|I+Q^{-1/2}D_V^{-1}Q^{-1/2}\big|^{-1/2} \le 1,$$

我们得到关于 V 边缘分布的如下界:

(20)
$$\begin{aligned}\int \pi(V\mid M,Y)\pi(M\mid V,Y)\,dM &\le C\Big(\prod_{i=1}^{n} V_i^{\tilde{p}-1}(b^{\eta}+V_i^{\eta})\big((1+b)^{\bar{\eta}}+V_i^{\bar{\eta}}\big)\Big)S_{\eta}^{n/2}S_{\bar{\eta}}^{n/2}\\ &\quad\times\exp\Big\{-\tfrac{1}{2}b\,\mathbf{1}^{\top}D_V^{-1}\mathbf{1}\Big\}\sum_{\phi\in\Phi}\exp\big\{-\tfrac{1}{2}\Delta_{\phi}(V)\big\}\\ &\le C\max\left\{\left(\prod_{i=1}^{n}V_i\right)^{m_1},\left(\prod_{i=1}^{n}V_i^{-1}\right)^{m_2}\right\},\end{aligned}$$

其中 $C>0$ 是一个常数， m_1， $m_2\in\mathbb{N}$，并且指数间隙Δ_{ϕ}（V）定义为

$$\Delta_{\phi}(V):=\tilde{a}\,\mathbf{1}^{\top}D_V\mathbf{1}+m^{\top}Q^{-1}m-(m+\sqrt{\tilde{a}}\phi)^{\top}(Q+D_V^{-1})^{-1}(m+\sqrt{\tilde{a}}\phi).$$

我们证明，对于所有 $\phi\in\Phi$,

(21)
$$\Delta_{\phi}(V)\;\ge\;\frac{3}{8}(\sqrt{\tilde{a}}-|\mu|)^2\sum_{i=1}^{n}V_i\;-\;\frac{1}{2}\|x\|_2^2.$$

将 $A=D_V$、 $B=Q^{-1}$、 $u=\sqrt{\tilde{a}}\phi$ 和 $v=m$ 代入引理 4.1，可得:

$$\Delta_{\phi}(V)=\big(\sqrt{\tilde{a}}D_V\phi-Q^{-1}m\big)^{\top}(D_V+Q^{-1})^{-1}\big(\sqrt{\tilde{a}}D_V\phi-Q^{-1}m\big).$$

由于 $Q\succeq D_V^{-1}$、 $Q^{-1}\preceq D_V$。因此， $D_V+Q^{-1}\preceq 2D_V$，从而（D_V+Q^{-1}）$^{-1}\succeq\frac{1}{2}D_V^{-1}$。所以,

(22)
$$\Delta_{\phi}(V)\ge\tfrac{1}{2}\big(\sqrt{\tilde{a}}D_V\phi-Q^{-1}m\big)^{\top}D_V^{-1}\big(\sqrt{\tilde{a}}D_V\phi-Q^{-1}m\big).$$

代入 $m=B^{\top}x+\mu\mathbf{1}$ 并定义 $z:=\sqrt{\tilde{a}}D_V\phi-\bar{}\mu Q^{-1}\mathbf{1}$，可得

$$\sqrt{\tilde{a}}D_V\phi-Q^{-1}m=z-Q^{-1}B^{\top}x.$$

于是，方程（22）变为:

(23)
$$\Delta_{\phi}(V)\ge\tfrac{1}{2}(z-Q^{-1}B^{\top}x)^{\top}D_V^{-1}(z-Q^{-1}B^{\top}x).$$

展开该表达式并应用不等式 $2u^{\top}v\le u^{\top}u+v^{\top}v$，其中 $u=x$ 且 $v=BQ^{-1}D_V^{-1}z$, 可得

$$\begin{aligned}&(z-Q^{-1}B^{\top}x)^{\top}D_V^{-1}(z-Q^{-1}B^{\top}x)\\ &\qquad\ge z^{\top}\Big(D_V^{-1}-D_V^{-1}Q^{-1}D_V^{-1}+D_V^{-1}Q^{-1}D_V^{-1}Q^{-1}D_V^{-1}\Big)z-\|x\|_2^2,\end{aligned}$$

其中我们使用了关系式 $B^{\top}B=Q-D_V^{-1}$，并舍弃了非负项$x^{\top}BQ^{-1}D_V^{-1}Q^{-1}B^{\top}x$。为简化记号，令 $D:=D_V$，并定义对称正定矩阵 $G:=D^{1/2}QD^{1/2}\succ 0$。于是 $G^{-1}=D^{-1/2}Q^{-1}D^{-1/2}$ 且 $G^{-2}=D^{-1/2}Q^{-1}D^{-1}Q^{-1}D^{-1/2}$。直接进行因式分解可知，括号中的矩阵可改写为

$$D^{-1}-D^{-1}Q^{-1}D^{-1}+D^{-1}Q^{-1}D^{-1}Q^{-1}D^{-1}=D^{-1/2}\big(I-G^{-1}+G^{-2}\big)D^{-1/2}.$$

接下来，注意到对所有 $t>0$，标量不等式

$$1-\frac{1}{t}+\frac{1}{t^2}\;\ge\;\frac{3}{4}$$

成立，这等价于 $\tfrac{1}{4}t+t^{-1}\ge 1$。由谱函数演算，这对所有 $G\succ 0$ 给出矩阵不等式 $I-G^{-1}+G^{-2}\succeq \tfrac{3}{4}I$。因此，我们有：

$$z^\top\Big(D^{-1}-D^{-1}Q^{-1}D^{-1}+D^{-1}Q^{-1}D^{-1}Q^{-1}D^{-1}\Big)z \\ =z^\top D^{-1/2}\big(I-G^{-1}+G^{-2}\big)D^{-1/2}z\ge \tfrac{3}{4}\,z^\top D^{-1}z.$$

由此得到以下界：

$$(z-Q^{-1}B^\top x)^\top D^{-1}(z-Q^{-1}B^\top x)\ge \tfrac{3}{4}\,z^\top D^{-1}z-\|x\|_2^2.$$

将（23）与上述不等式结合，可得：

$$\Delta_\phi(V)\ge \frac{1}{2}\Big(\tfrac{3}{4}\,z^\top D_V^{-1}z-\|x\|_2^2\Big)=\frac{3}{8}\,z^\top D_V^{-1}z-\frac{1}{2}\|x\|_2^2. \tag{24}$$

最后，我们对 $z^\top D_V^{-1}z$ 进行界定。由于 $\phi_i^2=1$，可得：

$$z^\top D_V^{-1}z=\sum_{i=1}^n\frac{(\sqrt{\tilde a}V_i\phi_i-\mu(Q^{-1}\mathbf{1})_i)^2}{V_i}.$$

应用加权柯西–施瓦茨不等式，

$$\sum_{i=1}^n\frac{r_i^2}{V_i}\ge\frac{\big(\sum_{i=1}^n r_i\big)^2}{\sum_{i=1}^n V_i},$$

其中 $r_i=\sqrt{\tilde a}V_i-\mu\,\phi_i\,(Q^{-1}\mathbf{1})_i$，可得：

$$z^\top D_V^{-1}z\ge\frac{\big(\sqrt{\tilde a}\sum_i V_i-\mu\,\phi^\top Q^{-1}\mathbf{1}\big)^2}{\sum_i V_i}.$$

此外，利用 $Q^{-1}\preceq D_V$ 以及带有 $u=Q^{-1/2}\phi$ 和 $v=Q^{-1/2}\mathbf{1}$的不等式 $2|u^\top v|\le u^\top u+v^\top v$，可得：

$$|\phi^\top Q^{-1}\mathbf{1}|\le\tfrac{1}{2}\phi^\top Q^{-1}\phi+\tfrac{1}{2}\mathbf{1}^\top Q^{-1}\mathbf{1}\le\tfrac{1}{2}\phi^\top D_V\phi+\tfrac{1}{2}\mathbf{1}^\top D_V\mathbf{1}=\sum_{i=1}^n V_i.$$

这表明$\big|\sqrt{\tilde a}\sum_i V_i-\mu\phi^\top Q^{-1}\mathbf{1}\big|\ge(\sqrt{\tilde a}-|\mu|)\sum_i V_i$，因此我们有$z^\top D_V^{-1}z\ge(\sqrt{\tilde a}-|\mu|)^2\sum_{i=1}^n V_i$。将此结果代入(24)即可证明(21)。

由(21)，我们关于 $\phi\in\Phi$一致地有：

$$\exp\big\{-\tfrac{1}{2}\Delta_\phi(V)\big\}\le\exp\big\{\tfrac{1}{4}\|x\|_2^2\big\}\exp\Big\{-c\sum_{i=1}^n V_i\Big\},\qquad c:=\tfrac{3}{16}(\sqrt{\tilde a}-|\mu|)^2>0.$$

注意到$\exp\{-\frac{1}{2}b\,\mathbf{1}^\top D_V^{-1}\mathbf{1}\}=\exp\{-\frac{b}{2}\sum_i V_i^{-1}\}$，并将$\exp\{\|x\|_2^2/4\}$和$|\Phi|=2^n$ 吸收到 C 中，我们由(20)得到：

$$\int\pi(V\mid M,Y)\pi(M\mid V,Y)\,dM\le C\max\left\{\Big(\prod_{i=1}^n V_i\Big)^{m_1},\Big(\prod_{i=1}^n V_i^{-1}\Big)^{m_2}\right\} \\ \times\exp\left\{-c\sum_{i=1}^n V_i-\frac{b}{2}\sum_{i=1}^n V_i^{-1}\right\}.$$

由于 $c>0$，因子 指数函数$\{-c\sum_i V_i\}$ 会随任意 $V_i\to\infty$ 指数衰减，从而压过 $(\prod_i V_i)^{m_1}$至多呈多项式增长的趋势。类似地，由于 $b>0$，因子

指数函数$\{-\frac{b}{2}\sum_i V_i^{-1}\}$会随任意 $V_i \to 0$ 指数衰减，从而压过 $(\prod_i V_i^{-1})^{m_2}$至多呈多项式增长的趋势。因此，右端在 $(0,\infty)^n$ 上可积，由此得出

$$\int_{(0,\infty)^n}\int_{\mathbb{R}^n} \pi(V \mid M,Y)\pi(M \mid V,Y)\,dM\,dV < \infty,$$

由此确立了情形 1 中的迹类性质。

情形 *2*： $a>0$, $b=0$, $p > \frac{1}{2}$。论证直至 (20) 的过程完全相同，只是指数项 $\exp\{-\frac{b}{2}\sum_i V_i^{-1}\}$ 不存在（因为 $b=0$）。当$V_i \to \infty$ 时的指数衰减仍然成立，因为 $c>0$ 依然成立（事实上，$a>0$ 蕴含 $\sqrt{\tilde{a}} > |\mu|$）。因此，我们只需验证在$V_i \to 0$ 附近的可积性。

对于 $p > \frac{1}{2}$，我们有 $\tilde{p} = p - \frac{1}{2} > 0$，因而 $\eta = \frac{1}{2}(|\tilde{p}| - \tilde{p}) = 0$。因此，(20) 中前置因子在 V 较小时的主导行为为

$$\prod_{i=1}^{n} V_i^{\tilde{p}-1} = \prod_{i=1}^{n} V_i^{p-\frac{3}{2}},$$

直至当 $V_i \to 0$ 时仍保持有界的乘性因子（因为当 $\eta = 0$ 时 $S_\eta \equiv 1$，且 $\bar{\eta} \geq 0$）。因此，在 0 附近的 V-积分有限当且仅当 $p - \frac{3}{2} > -1$，这等价于 $p > \frac{1}{2}$。这与所述假设相符，因而目标积分在情形 2 中有限。至此完成了情形 (1)–(2) 中迹类性质的证明。□

4.2. 几何遍历性证明。 我们通过验证漂移条件和一步小集条件，并随后应用 Rosenthal 漂移-小集定理 [30,定理 12]，来证明边际 Gibbs 链 $(V^{(t)})_{t\geq 0}$ 具有几何遍历性。为完整起见，我们在下文重述该定理。

定理 4.1（Rosenthal 漂移-小集界 [30, 定理 12]）。 设 $P(x,\cdot)$ 为马尔可夫链 $(X^{(k)})_{k\geq 0}$ 在状态空间 $\mathcal{X}$ 上的转移核，其平稳分布为 π。假设存在可测函数 $V:\mathcal{X} \to \mathbb{R}_{\geq 0}$、常数 $\lambda \in (0,1)$ 和$b < \infty$，使得漂移条件成立：

$$\mathbb{E}\Big[V(X^{(1)}) \mid X^{(0)} = x\Big] \;\leq\; \lambda V(x) + b, \qquad \forall x \in \mathcal{X}.$$

进一步假设存在 $\varepsilon > 0$、定义在 $\mathcal{X}$ 上的概率测度 $Q(\cdot)$，以及 $d > 2b/(1-\lambda)$，使得小化条件成立：

$$P(x,\cdot) \;\geq\; \varepsilon\, Q(\cdot), \qquad \forall x \in \mathcal{X} \text{ with } V(x) \leq d.$$

定义

$$\alpha^{-1} := \frac{1 + 2b + \lambda d}{1+d} \in (0,1), \qquad A := 1 + 2(\lambda d + b).$$

那么，对于任意 $0 < r < 1$ 以及任意初始分布 ν，

$$\big\|\mathcal{L}(X^{(k)}) - \pi\big\|_{\mathrm{TV}} \;\leq\; (1-\varepsilon)^{rk} \;+\; \Big(\alpha^{-(1-r)}A^{r}\Big)^{k}\Big(1 + \frac{b}{1-\lambda} + \mathbb{E}_{\nu}[V(X^{(0)})]\Big).$$

特别地，该链具有几何遍历性。

该定理给出了吉布斯采样器具有几何遍历性的充分条件。现在，我们为特定模型建立漂移与小化条件。

对于（ 5 $\rho=\sigma^2\sigma_\epsilon^{-2}$ $\bar{Q}=\sigma^2Q=\rho B^\top B+D_V^{-1}$ $\bar{m}=\sigma^2m=$）中的模型，我 $\rho B^\top$ $Y-X\beta+\mu Bh$ $+\mu\mathbf{1}$
们记为，,以及 ()。吉布斯采样器具有以下转移分布：

(25a) $$M|V,Y\sim N(\bar{Q}^{-1}\bar{m},\sigma^2\bar{Q}^{-1})$$

(25b) $$\tilde{V}_i|M_i,Y\overset{iid}{\sim}\text{GIG}\left(p-\frac{1}{2},a+\left(\frac{\mu}{\sigma}\right)^2,b+\left(\frac{M_i}{\sigma}\right)^2\right)$$

我们通过考虑以下三种情形来建立漂移条件：

情形 1： $a=0,\ p<0,\ b>0,\ \mu=0$;
情形 2： $a=0,\ p<0,\ b>0,\ \mu\neq0$;
情形 3： $a>0,\ b=0,\ p>0$。

4.2.1. 情形 *1*。我们首先考虑特殊情形 $a=\mu=0$，这意味着 $b>0$ 和$p<0$。在此情形下，式 (25) 中的转移分布简化为

(26a) $$M|V,Y\sim N(\bar{Q}^{-1}\bar{m},\sigma^2\bar{Q}^{-1})$$

(26b) $$\tilde{V}_i|M_i,Y\overset{iid}{\sim}\text{Inv-Gamma}\left(-p+\frac{1}{2},\frac{b}{2}+\frac{M_i^2}{2\sigma^2}\right)$$

回忆一下，形状参数为 $\alpha>0$、尺度参数为$\beta>0$ 的逆伽马分布 Inv-Gamma$(\alpha,\ \beta)$ 的密度为

$$f_X(x)=\frac{\beta^\alpha}{\Gamma(\alpha)}x^{-(\alpha+1)}\exp\left(-\frac{\beta}{x}\right),\quad x>0.$$

在此情形下，可以证明马尔可夫算子通常不是迹类。我们在以下引理中正式陈述这一点（证明见附录 B）。

引理 4.4. 假设 $a=\mu=0$ （蕴含 $p<0$ 与 $b>0$），以及 $A=0$。则与吉布斯采样器相关联的马尔可夫算子不是迹类。

备注。情形 $\mu=0$ 满足零空间小性假设（假设 2.1）。因此，引理 4.4 表明仅凭该假设并不足以保证迹类性质。然而，如下文所示，它足以保证几何遍历性。

尽管迹类性质不成立，我们仍可通过验证漂移与小化条件来建立几何遍历性。我们首先引入若干技术引理（第一个的证明推迟至附录 B）。

引理 4.5（一致有界性： $\eta(V)$）。令 $\bar{m}\in$ 值域$(B^\top)$。存在常数 $C<\infty$，使得

$$\left\|\left(\rho B^\top B+D_V^{-1}\right)^{-1}\bar{m}\right\|_2\leq C,$$

对所有 $V_1,\ldots,V_n\in\in(0,\infty)$，其中 $D_V==\text{diag}(V_1,\ldots,V_n)$。

引理 4.6. 令 $\alpha>0$ 和 $\delta\in(0,\ \alpha)$。则

(27) $$\frac{\Gamma(\alpha+\frac{1}{2}-\delta)}{\Gamma(\alpha+\frac{1}{2})}\cdot\frac{\Gamma(\delta+\frac{1}{2})}{\sqrt{\pi}}<1.$$

证明。定义

$$f(\delta) := \log\Gamma\big(\alpha+\tfrac{1}{2}-\delta\big) + \log\Gamma\big(\delta+\tfrac{1}{2}\big) - \log\Gamma\big(\alpha+\tfrac{1}{2}\big) - \log\Gamma\big(\tfrac{1}{2}\big).$$

于是 $f''(\delta) = \psi_1\big(\alpha+\tfrac{1}{2}-\delta\big) + \psi_1\big(\delta+\tfrac{1}{2}\big) > 0$，因此 f在$(0,\alpha)$上严格凸。由于$f(0) = 0$且 $f(\alpha) = 0$，严格凸性意味着对所有 $\delta \in (0,\alpha)$，都有 $f(\delta) < 0$。取指数即可得到（27）。 □

为在 $a = \mu = 0$ 的情形下建立漂移条件，我们引入以下记号。回忆一下，在假设 $a = \mu = 0$ （其中 $b > 0$ 和 $p < 0$）下，分布简化为式（ 26 ）中的逆伽马族。令 $\alpha := -p > 0$，并定义 $\delta :=$ 最小值 $\{\alpha/2, 1/2\} \in (0, 1/2]$。我们使用李雅普诺夫函数

$$G(V) := 1 + \sum_{i=1}^{n} \Big(V_i^{\delta} + V_i^{-1/4} \Big).$$

回顾一下， $\rho := \sigma^2\sigma_\epsilon^{-2}$， $\bar{Q}(V) := \rho B^\top B + D_V^{-1}$，且 $\bar{m} := \rho B^\top (Y - X\boldsymbol{\beta}) \in$ 值域$(B^\top)$。给定 $(V,\ Y)$ 时， M的条件分布为

$$M \mid (V,Y) \sim N\big(\eta(V), \Sigma(V)\big), \qquad \eta(V) := \bar{Q}(V)^{-1}\bar{m}, \quad \Sigma(V) := \sigma^2\bar{Q}(V)^{-1},$$

且给定 $M_i = x$ 时， V_i 的条件分布为

$$\tilde{V}_i \mid (M_i = x) \sim \text{Inv-Gamma}\Big(\alpha+\tfrac{1}{2},\ \frac{b}{2} + \frac{x^2}{2\sigma^2}\Big).$$

定义常数

$$C_\delta := \frac{\Gamma(\alpha+\frac{1}{2}-\delta)}{\Gamma(\alpha+\frac{1}{2})}, \qquad K_\delta := \frac{\Gamma(\delta+\frac{1}{2})}{\sqrt{\pi}}, \qquad \gamma^{(\delta)} := C_\delta\, K_\delta,$$

and

$$\phi_{\mathrm{IG}}^{(\delta)} := 1 + nC_\delta\Big(\frac{b}{2}\Big)^{\delta} + n\,\frac{\Gamma(\alpha+\frac{3}{4})}{\Gamma(\alpha+\frac{1}{2})}\Big(\frac{b}{2}\Big)^{-1/4}.$$

令 C_η 表示引理 4.5 中的常数，使得 $\sup_{V\in(0,\infty)^n} \|\eta(V)\|_2 \le C_\eta$，并令

$$\psi^{(\delta)} := \frac{C_\delta}{(2\sigma^2)^\delta}\, n^{1-\delta} C_\eta^{2\delta}.$$

引理 4.7（关于 $a = \mu = 0$ （逆伽马分布）的漂移条件）。在上述假设和记号下，对所有 $V \in (0,\ \infty)^n$,

$$\mathbb{E}\Big[G(\tilde{V}) \mid V\Big] \le \gamma^{(\delta)}\, G(V) + \big(\phi_{\mathrm{IG}}^{(\delta)} + \psi^{(\delta)}\big).$$

此外， $\gamma^{(\delta)} < 1$ 根据引理 *4.6*，因为 $\delta \in (0,\ \alpha)$，且 $\alpha = -p$。

证明。对于逆伽马分布 Inv-Gamma$(\alpha+\frac{1}{2}, \frac{b}{2}+\frac{x^2}{2\sigma^2})$，分数阶矩公式对任意 $r < \alpha + \frac{1}{2}$ （可以为负数）均成立： –

$$\mathbb{E}\big(\tilde{V}_i^{\,r} \mid M_i = x, Y\big) = \frac{\Gamma(\alpha+\frac{1}{2}-r)}{\Gamma(\alpha+\frac{1}{2})}\Big(\frac{b}{2}+\frac{x^2}{2\sigma^2}\Big)^r.$$

令 $r = \delta$，并对 $\delta \in (u+v)^\delta \le u^\delta + v^\delta$ 使用不等式$(0,1]$ 和 $u, v \ge 0$，可得

$$\mathbb{E}\big(\tilde{V}_i^{\,\delta} \mid M_i = x, Y\big) \le C_\delta\Big(\frac{b}{2}\Big)^{\delta} + C_\delta\,\frac{|x|^{2\delta}}{(2\sigma^2)^\delta}. \tag{28}$$

对于负阶矩，我们对 $r=-\frac{1}{4}$应用相同的公式：

$$\mathbb{E}\big(\tilde{V}_i^{-1/4}\mid M_i=x,Y\big)=\frac{\Gamma(\alpha+\frac{3}{4})}{\Gamma(\alpha+\frac{1}{2})}\Big(\frac{b}{2}+\frac{x^2}{2\sigma^2}\Big)^{-1/4}\le\frac{\Gamma(\alpha+\frac{3}{4})}{\Gamma(\alpha+\frac{1}{2})}\Big(\frac{b}{2}\Big)^{-1/4}. \tag{29}$$

˜根据转移核的定义，给定(M,V,Y)˜ ˜时， V的条件分布与给定(M,Y)时， V的条件分布相同。换言之，给定(M,Y)时， V与 V条件独立。因此，对于任意可测函数 h，有$\mathbb{E}\big[h(\tilde{V})\mid M,Y\big]=\mathbb{E}\big[h(\tilde{V})\mid M,V,Y\big]$。根据塔式性质，

$$\mathbb{E}\Big[\mathbb{E}\big[h(\tilde{V})\mid M,Y\big]\,\Big|\,V\Big]=\mathbb{E}\Big[\mathbb{E}\big[h(\tilde{V})\mid M,V,Y\big]\,\Big|\,V\Big]=\mathbb{E}\big[h(\tilde{V})\mid V\big].$$

将此恒等式应用于 $h(\tilde{V})=\sum_{i=1}^n\tilde{V}_i^{\delta}$和 $h(\tilde{V})=\sum_{i=1}^n\tilde{V}_i^{-1/4}$，并对 i求和，将界（28）和（29）相加，可得

$$\mathbb{E}\left[\sum_{i=1}^n\tilde{V}_i^{\delta}\mid V\right]\le nC_\delta\Big(\frac{b}{2}\Big)^{\delta}+\frac{C_\delta}{(2\sigma^2)^{\delta}}\,\mathbb{E}\left[\sum_{i=1}^n|M_i|^{2\delta}\mid V\right], \tag{30}$$

$$\mathbb{E}\left[\sum_{i=1}^n\tilde{V}_i^{-1/4}\mid V\right]\le n\,\frac{\Gamma(\alpha+\frac{3}{4})}{\Gamma(\alpha+\frac{1}{2})}\Big(\frac{b}{2}\Big)^{-1/4}. \tag{31}$$

在给定 V 的条件下，写作 $M_i=\eta_i(V)+Z_i$，其中 $Z_i\sim N(0,\Sigma_{ii}(V))$。由于 $\delta\le\frac{1}{2}$，我们有 $2\delta\le1$，因此 $|a+b|^{2\delta}\le|a|^{2\delta}+|b|^{2\delta}$。所以

$$\mathbb{E}\big(|M_i|^{2\delta}\mid V\big)\le|\eta_i(V)|^{2\delta}+\mathbb{E}|Z_i|^{2\delta}.$$

令 $m_{2\delta}:=\mathbb{E}|Z|^{2\delta}=2^{\delta}\Gamma(\delta+\frac{1}{2})/\sqrt{\pi}$，其中 $Z\sim N(0,1)$。由于 $Z_i=\Sigma_{ii}(V)^{1/2}Z$，我们有$\mathbb{E}|Z_i|^{2\delta}=m_{2\delta}\,\Sigma_{ii}(V)^{\delta}$。对 i 求和可得

$$\mathbb{E}\left[\sum_{i=1}^n|M_i|^{2\delta}\mid V\right]\le\sum_{i=1}^n|\eta_i(V)|^{2\delta}+m_{2\delta}\sum_{i=1}^n\Sigma_{ii}(V)^{\delta}. \tag{32}$$

现在我们分别处理右侧的两项（32）。首先，考虑方差项。由于 $\bar{Q}(V)$ $=\rho B^{\top}B+D_V^{-1}\succeq D_V^{-1}$，我们有 $\bar{Q}(V)^{-1}\preceq D_V$。因此 $(\bar{Q}(V)^{-1})_{ii}\le V_i$，从而

$$\Sigma_{ii}(V)=\sigma^2(\bar{Q}(V)^{-1})_{ii}\le\sigma^2V_i,\qquad\sum_{i=1}^n\Sigma_{ii}(V)^{\delta}\le\sigma^{2\delta}\sum_{i=1}^nV_i^{\delta}.$$

接下来，我们对均值项进行界定。根据引理 4.5，$\sup_V\|\eta(V)\|_2\le C_\eta$。因此

$$\sum_{i=1}^n|\eta_i(V)|^{2\delta}=\|\eta(V)\|_{2\delta}^{2\delta}\le\big(n^{\frac{1}{2\delta}-\frac{1}{2}}\|\eta(V)\|_2\big)^{2\delta}=n^{1-\delta}\|\eta(V)\|_2^{2\delta}\le n^{1-\delta}C_\eta^{2\delta}.$$

将这两个界代入 (32) 可得

$$\mathbb{E}\left[\sum_{i=1}^n|M_i|^{2\delta}\mid V\right]\le n^{1-\delta}C_\eta^{2\delta}+m_{2\delta}\sigma^{2\delta}\sum_{i=1}^nV_i^{\delta}.$$

代入式（30）：

$$\mathbb{E}\left[\sum_{i=1}^n\tilde{V}_i^{\delta}\mid V\right]\le nC_\delta\Big(\frac{b}{2}\Big)^{\delta}+\frac{C_\delta}{(2\sigma^2)^{\delta}}n^{1-\delta}C_\eta^{2\delta}+\frac{C_\delta}{(2\sigma^2)^{\delta}}m_{2\delta}\sigma^{2\delta}\sum_{i=1}^nV_i^{\delta}.$$

注意，$m_{2\delta}/2^{\delta} = \Gamma\left(\delta + \tfrac{1}{2}\right)/\sqrt{\pi} = K_{\delta}$，因此最后一个系数等于 $C_{\delta}K_{\delta} = \gamma^{(\delta)}$。因此

$$\mathbb{E}\left[\sum_{i=1}^{n} \tilde{V}_i^{\delta} \mid V\right] \le nC_{\delta}\left(\frac{b}{2}\right)^{\delta} + \psi^{(\delta)} + \gamma^{(\delta)} \sum_{i=1}^{n} V_i^{\delta}.$$

最后，结合式 (31) 并利用 $\sum_i V_i^{\delta} \le G(V)$，可得

$$\mathbb{E}[G(\tilde{V}) \mid V] \le \gamma^{(\delta)}\, G(V) + \left(\phi_{\mathrm{IG}}^{(\delta)} + \psi^{(\delta)}\right).$$

根据引理 4.6，$\gamma^{(\delta)} < 1$。 □

4.2.2. 情形 *2*。 对于这种情形，其中 $\tilde{a} := \mu^2/\sigma^2 > 0$，我们需要假设 2.1。为建立漂移条件，我们需要若干技术引理。第一个引理给出了将全 1 向量投影到 B 的零空间上的计算公式，这对于界定 $\bar{m}$ 在零空间中的分量至关重要。

引理 4.8。 假设 $B = AK^{-1}$ 且 $K \in \mathbb{R}^{n\times n}$ 可逆。设 $U_A \in \mathbb{R}^{n\times r}$ 是 零空间(A) 的一组标准正交基，其中 $r =$零空间(A) 的维数为 $r =$维数零空间(A)。定义 $G := U_A^{\top} K^{\top} K U_A \in \mathbb{R}^{r\times r}$，以及 $z := U_A^{\top} K^{\top} \mathbf{1} \in \mathbb{R}^r$。令 $P_{\mathrm{Null}(B)}$ 表示到 零空间(B) 上的欧几里得正交投影算子。则

$$\|P_{\mathrm{Null}(B)}\mathbf{1}\|_2^2 = \mathbf{1}^{\top} P_{\mathrm{Null}(B)}\mathbf{1} = z^{\top} G^{-1} z. \tag{33}$$

特别地，如果 $\bar{m} = \rho B^{\top}(Y - X\beta + \mu Bh) + \mu\mathbf{1}$ 且 $\bar{m}_0 := P_{\mathrm{Null}(B)}\bar{m}$，那么我们有 $\bar{m}_0 = \mu P_{\mathrm{Null}(B)}\mathbf{1}$，因此

$$\|\bar{m}_0\|_2 = |\mu|\, \|P_{\mathrm{Null}(B)}\mathbf{1}\|_2 = |\mu|\sqrt{z^{\top} G^{-1} z}. \tag{34}$$

第二个引理给出了广义逆高斯分布负阶矩的一致有界性，这对于在参数变化时控制 V_i 的行为是必要的。

引理 4.9. 假设 $p < 0$、 $\tilde{a} > 0$ 和 $b > 0$。任取 $q \in (0, \tfrac{1}{2}]$。则存在一个常数 $L_{-}(p,\ \tilde{a},\ b,\ q) < \infty$，使得对于所有 $M_i \in \mathbb{R}$、 $\mathbb{E}\left[\tilde{V}_i^{-q} \mid M_i\right] \le L_{-}(p,\ \tilde{a},\ b,\ q)$。因此，对于所有 $V \in (0,\ \infty)^n$，

$$\sum_{i=1}^{n} \mathbb{E}\left[\tilde{V}_i^{-q} \mid V\right] \le n\, L_{-}(p, \tilde{a}, b, q).$$

最后，第三个引理给出了广义逆高斯分布一阶矩的线性上界，该上界对形状参数 p 的所有取值均一致成立。这对于˜在漂移分析中界定涉及 $\mathbb{E}[V_i \mid M_i]$ 的项至关重要。

引理 4.10. 设 $\tilde{V}_i \mid M_i \sim$广义逆高斯分布$(p - \tfrac{1}{2}, \tilde{a},\ b + (M_i/\sigma)^2)$ 满足 $\tilde{a} > 0$ 且 $b \ge 0$。则存在常数 $C_{\mathrm{lin}}(p, \tilde{a}) < \infty$，使得对所有 $M_i \in \mathbb{R}$，

$$\mathbb{E}[\tilde{V}_i \mid M_i] \le \frac{1}{\sqrt{\tilde{a}}}\left|\frac{M_i}{\sigma}\right| + \frac{\sqrt{b}}{\sqrt{\tilde{a}}} + \frac{C_{\mathrm{lin}}(p, \tilde{a})}{\tilde{a}}.$$

引理 4.8、4.9 和 4.10 的证明见附录 B。有了这些辅助引理，我们现在可以为情形 2 建立主要的漂移条件。

引理 4.11（$a=0,\ \mu\neq 0$的漂移条件）。假设 $a=0$、 $\mu\neq 0$、 $b>0$ 和 $p<0$。设假设 *2.1* 成立。任取 $q\in(0,\frac{1}{2}]$，并定义李雅普诺夫函数

$$G_q(V) := 1+\sum_{i=1}^{n}\Big(V_i+V_i^{-q}\Big).$$

于是存在 $\gamma\in(0,1)$ 和 $L<\infty$，使得对所有 $V\in(0,\infty)^n$，

$$\mathbb{E}\Big[G_q(\tilde V)\mid V\Big]\ \le\ \gamma\,G_q(V)+L.$$

*M*更确切地说，对于任何满足 $\gamma_+ := \sqrt{z^\top G^{-1}\overline{z+\varepsilon<1}}$ 的 $\varepsilon>0$，可以取 $\gamma=\gamma_+$.

证明。令 $\eta(V):=\bar Q(V)^{-1}\bar m$，并回顾 $\bar Q(V)=\rho B^\top B+D_V^{-1}$。由条件期望的线性性，

$$\mathbb{E}\Big[G_q(\tilde V)\mid V\Big]=1+\sum_{i=1}^{n}\mathbb{E}[\tilde V_i\mid V]+\sum_{i=1}^{n}\mathbb{E}[\tilde V_i^{-q}\mid V].$$

由引理 4.9，负阶矩一致有界：

$$\sum_{i=1}^{n}\mathbb{E}[\tilde V_i^{-q}\mid V]\le n\,L_-(p,\tilde a,b,q)=:L_-.$$

由引理 4.10，

$$\sum_{i=1}^{n}\mathbb{E}[\tilde V_i\mid V]\le\frac{1}{\sqrt{\tilde a}}\,\mathbb{E}\Big[\Big\|\frac{M}{\sigma}\Big\|_1\,\Big|\,V\Big]+n\Big(\frac{\sqrt{b}}{\sqrt{\tilde a}}+\frac{C_{\mathrm{lin}}(p,\tilde a)}{\tilde a}\Big).$$

回顾一下， $M/\sigma\mid V\sim N(\eta(V)/\sigma,\,\bar Q(V)^{-1})$。利用这个条件分布以及三角不等式，我们得到

$$\mathbb{E}\left[\Big\|\frac{M}{\sigma}\Big\|_1\,\Big|\,V\right]\le\Big\|\frac{\eta(V)}{\sigma}\Big\|_1+\mathbb{E}\|Z\|_1,\qquad Z\sim N(0,\bar Q(V)^{-1}).$$

现在，将 $\bar m=\bar m_0+\bar m_\parallel$ 分解为 $\bar m_0:=P_{\mathrm{Null}(B)}\bar m$ 和 $\bar m_\parallel\in$ 值域$(B^\top)$。因此 $\eta(V)=\bar Q(V)^{-1}\bar m_0+\bar Q(V)^{-1}\bar m_\parallel$。对于零空间分量，应用引理 4.14 和 $\bar Q(V)^{-1}\preceq D_V$，得到

$$\|\bar Q(V)^{-1}\bar m_0\|_1\le\sqrt{\sum_{i=1}^{n}V_i}\,\sqrt{\bar m_0^\top\bar Q(V)^{-1}\bar m_0}\le\sqrt{\sum_{i=1}^{n}V_i}\,\sqrt{\bar m_0^\top D_V\bar m_0}\le\|\bar m_0\|_2\sum_{i=1}^{n}V_i.$$

现在应用引理 4.8。由于 $\bar m=\rho B^\top(Y-X\beta+\mu Bh)+\mu\mathbf{1}$，它在零空间上的投影完全由 μ决定： $\bar m_0=\mu P_{\mathrm{Null}(B)}\mathbf{1}$。根据引理 4.8（式（34））， 我们有 $\|\bar m_0\|_2=|\mu|\sqrt{z^\top G^{-1}z}$。将其代入零空间分量的界中， $\sum V_i$ 的系数变为：

$$\frac{1}{\sigma\sqrt{\tilde a}}\|\bar m_0\|_2=\frac{1}{\sigma(\mu/\sigma)}|\mu|\sqrt{z^\top G^{-1}z}=\sqrt{z^\top G^{-1}z}.$$

对于值域分量，正如引理 4.15（步骤 2(A)）所证明的，存在 $c_\parallel<\infty$使得

$$\frac{1}{\sqrt{\tilde a}}\Big\|\frac{\bar Q(V)^{-1}\bar m_\parallel}{\sigma}\Big\|_1\le c_\parallel\sqrt{\sum V_i}\le\frac{\varepsilon}{2}\sum V_i+\frac{c_\parallel^2}{2\varepsilon}.$$

类似地，波动项满足（其中 $c_0 := \sqrt{2/\pi}/\sqrt{\tilde{a}}$）：

$$\frac{1}{\sqrt{\tilde{a}}}\mathbb{E}\|Z\|_1 \le c_0 \sum_{i=1}^{n} \sqrt{V_i} \le \frac{\varepsilon}{2}\sum_{i=1}^{n} V_i + \frac{n c_0^2}{2\varepsilon}.$$

结合这些界，存在 $L_+ < \infty$ 使得

$$\sum_{i=1}^{n} \mathbb{E}[\tilde{V}_i \mid V] \le \Big(\sqrt{z^\top G^{-1} z} + \varepsilon\Big) \sum_{i=1}^{n} V_i + L_+.$$

根据假设 2.1（情形 $a = 0$），我们有 $\sqrt{z^\top G^{-1} z} < 1$。因此，我们可以选择足够小的 $\varepsilon > 0$，使得 $\gamma_+ := \sqrt{z^\top G^{-1} z} + \varepsilon < 1$。令 $\gamma := \gamma_+$ 且 $L := 1 + L_- + L_+$。那么，对所有 $V \in (0,\ \infty)^n$，

$$\mathbb{E}\Big[G_q(\tilde{V}) \mid V\Big] \le \gamma \sum_{i=1}^{n} V_i + L \le \gamma\, G_q(V) + L,$$

这就证明了漂移条件。 □

4.2.3. 情形 *3*： $a > 0,\ b = 0,\ p > 0$。对于这种情形，证明需要几个技术引理。第一个引理给出了矩阵逆的对角元素的下界，以及高斯随机变量负阶矩的上界。

引理 4.12. 设 $A \succ 0$ 为对称矩阵。对于每个 i，$(A^{-1})_{ii} \ge 1/A_{ii}$。此外，对于任意 $0 < \delta < 1$ 和 $Z \sim N(\eta,\ s^2)$，

$$\mathbb{E}|Z|^{-\delta} \ \le\ \kappa(\delta)\, s^{-\delta}, \qquad \kappa(\delta) := \mathbb{E}|N(0,1)|^{-\delta} = \frac{1}{\sqrt{\pi}}\, 2^{-\delta/2}\Gamma\Big(\frac{1-\delta}{2}\Big).$$

证明。第一个不等式由舒尔补公式得出；等价地，也可由对角元素的变分表示得出。对于高斯界，注意到 $x \mapsto |x|^{-\delta}$ 关于 $|x|$ 对称且递减。由安德森不等式，$\mathbb{E}|N(\eta, s^2)|^{-\delta} \le \mathbb{E}|N(0, s^2)|^{-\delta} = s^{-\delta}\kappa(\delta)$。 □

第二个引理给出了广义逆高斯分布负分数阶矩的一个关键界，条件是 $p > 0$。该界涉及一个特定指数 $\delta(p)$，其选取旨在确保漂移系数具有压缩性。

引理 4.13. 给定 $\sigma > 0$ 和 $V_i \mid M_i \sim$ 广义逆高斯分布$(p - 1/2,\ \tilde{a},\ M_i^2/\sigma^2)$，其中 $\tilde{a} = a + \mu^2/\sigma^2 > 0$ 且 $p > 0$。定义

$$\delta(p) := \begin{cases} p & \text{if } 0 < p \le \frac{1}{2}, \\ \min\{\frac{1}{2}, 2p - 1\} & \text{if } p > \frac{1}{2}. \end{cases}$$

则 $0 < \delta(p) \le \frac{1}{2}$。存在常数 $C_1(p) > 0$（仅依赖于 p）以及 $C_2(p, \tilde{a}) \in \mathbb{R}$，使得

$$\mathbb{E}\left[V_i^{-\frac{\delta(p)}{2}} \mid M_i\right] \le C_1(p)\left|\frac{M_i}{\sigma}\right|^{-\delta(p)} + C_2(p, \tilde{a}).$$

此外，与引理 *4.12* 中的 $\kappa(\cdot)$ 相同，对于所有 $p > 0$，我们有 $C_1(p)\,\kappa\big(\delta(p)\big) < 1$。

第三个引理给出了一个 ℓ_1-对偶界，用于控制 $\bar{Q}(V)^{-1}u$ 的范数。

引理 4.14. 给定 $\bar{Q}(V) = \rho B^\top B + D_V^{-1}$，令 $u \in \mathbb{R}^n$ 和 $V \in (0, \infty)^n$。则

$$\|\bar{Q}(V)^{-1}u\|_1 \le \sqrt{\sum_{i=1}^n V_i}\sqrt{u^\top \bar{Q}(V)^{-1}u}. \tag{35}$$

引理 4.13 和 4.14 的证明见附录 B。在建立了这些技术性引理之后，现给出情形 3 的主要漂移条件。

引理 4.15（$a>0$的$b=0$漂移条件）。假设 $a>0$、$b=0$、$p>0$，以及假设 *2.1* 成立。令 $\delta = \delta(p) \in (0, \frac{1}{2}]$ 如引理 *4.13* 中所述，并定义李雅普诺夫函数

$$G_\delta(V) := 1 + \sum_{i=1}^n \Big(V_i + V_i^{-\delta/2}\Big).$$

则存在 $\gamma \in (0, 1)$ 以及 $L<\infty$（可能取决于维数 n），使得对于所有 $V \in (0, \infty)^n$、$\mathbb{E}\Big[G_\delta(\tilde{V}) \mid V\Big] \le \gamma G_\delta(V) + L$。明确地，对于满足以下条件的任意 $\varepsilon > 0$

$$\gamma_+ := \frac{|\mu|}{\sqrt{\sigma^2 a + \mu^2}}\sqrt{z^\top G^{-1} z} + \varepsilon < 1,$$

可取 $\gamma = \max\{\gamma_-, \gamma_+\} < 1$，其中 $\gamma_- := C_1(p)\kappa(\delta(p))$。

证明。令 $\mathcal{S} := \text{值域}(B^\top) \subseteq \mathbb{R}^n$，并令 $U \in \mathbb{R}^{n\times d}$ 为 $\mathcal{S}$ 的任意标准正交基矩阵（其中 $d = \text{维数 } \mathcal{S}$）。定义 $\lambda_+ := \lambda_{\min}\big(U^\top B^\top B U\big) > 0$，即 $B^\top B$ 的最小严格正特征值（等价地，为将 $B^\top B$ 限制在 值域$(B^\top)$ 上所得矩阵的最小特征值）。令 $\eta(V) := \bar{Q}(V)^{-1}\bar{m}$。由条件期望的线性性，

$$\mathbb{E}\Big[G_\delta(\tilde{V}) \mid V\Big] = 1 + \sum_{i=1}^n \mathbb{E}[\tilde{V}_i \mid V] + \sum_{i=1}^n \mathbb{E}[\tilde{V}_i^{-\delta/2} \mid V].$$

我们首先对负矩项进行界定。根据引理 4.13，对每个 i，

$$\mathbb{E}\Big[\tilde{V}_i^{-\delta/2} \mid M_i\Big] \le C_1(p)\Big|\frac{M_i}{\sigma}\Big|^{-\delta} + C_2(p, \tilde{a}).$$

在给定 V 的条件下，有 $Z_i := M_i/\sigma \sim N(\eta_i(V)/\sigma,\ (\bar{Q}(V)^{-1})_{ii})$。根据引理 4.12，

$$\mathbb{E}\Big[\big|Z_i\big|^{-\delta} \mid V\Big] \le \kappa(\delta)\,(\bar{Q}(V)^{-1})_{ii}^{-\delta/2}.$$

由于 $\bar{Q}(V)_{ii} = \rho(B^\top B)_{ii} + V_i^{-1}$ 且 $(\bar{Q}(V)^{-1})_{ii} \ge 1/\bar{Q}(V)_{ii}$，因此有不等式 $(\bar{Q}(V)^{-1})_{ii}^{-\delta/2} \le \bar{Q}(V)_{ii}^{\delta/2}$。利用 $(x+y)^q \le x^q + y^q$，对于 $q = \delta/2 \in (0, 1/4]$，我们得到

$$\bar{Q}(V)_{ii}^{\delta/2} \le \big(\rho(B^\top B)_{ii}\big)^{\delta/2} + V_i^{-\delta/2}.$$

结合这些估计可得 $\mathbb{E}[\tilde{V}_i^{-\delta/2} \mid V] \le C_1(p)\kappa(\delta)\,V_i^{-\delta/2} + L_{-,\,i}$，其中 $L_{-,\,i}$ 表示所有与 V 相互独立的项之和。对 i 求和并令 $L_- := \sum_{i=1}^n L_{-,\,i}$，我们得到

$$\sum_{i=1}^n \mathbb{E}[\tilde{V}_i^{-\delta/2} \mid V] \le \gamma_- \sum_{i=1}^n V_i^{-\delta/2} + L_-, \qquad \gamma_- := C_1(p)\kappa(\delta).$$

根据引理 4.13，$\gamma_- < 1$。接下来，我们对线性项进行界定。根据引理 4.10，

$$\sum_{i=1}^n \mathbb{E}[\tilde{V}_i \mid V] \le \frac{1}{\sqrt{\tilde{a}}}\,\mathbb{E}\Big[\Big\|\frac{M}{\sigma}\Big\|_1 \,\Big|\, V\Big] + n\Big(\frac{\sqrt{b}}{\sqrt{\tilde{a}}} + \frac{C_{\text{lin}}(p, \tilde{a})}{\tilde{a}}\Big).$$

由于在此情形下 $b=0$，且 $M/\sigma \mid V \sim N\left(\eta(V)/\sigma,\ \bar{Q}(V)^{-1}\right)$，我们可以将此条件分布与三角不等式结合，得到

$$\mathbb{E}\Big\|\frac{M}{\sigma}\Big\|_1 \,\Big|\, V \le \Big\|\frac{\eta(V)}{\sigma}\Big\|_1 + \mathbb{E}\|Z\|_1, \qquad \text{where } Z\sim N(0,\bar{Q}(V)^{-1}).$$

现在我们将 $\bar{m}=\bar{m}_0+\bar{m}_{\|}$ 分解，其中 $\bar{m}_0 := P_{\mathrm{Null}(B)}\bar{m}$ 且 $\bar{m}_{\|}\in$ 值域$(B^\top)$，使得 $\eta(V)=\bar{Q}(V)^{-1}\bar{m}_0+\bar{Q}(V)^{-1}\bar{m}_{\|}$。由引理 4.14 以及界 $\bar{Q}(V)^{-1}\preceq D_V$，

$$\Big\|\bar{Q}(V)^{-1}\bar{m}_0\Big\|_1 \le \sqrt{\sum_{i=1}^n V_i}\,\sqrt{\bar{m}_0^\top D_V \bar{m}_0} \le \|\bar{m}_0\|_2 \sum_{i=1}^n V_i.$$

引理 4.8 给出 $\|\bar{m}_0\|_2 = |\mu|\sqrt{z^\top G^{-1}z}$，因此

$$\frac{1}{\sigma\sqrt{\tilde{a}}}\Big\|\bar{Q}(V)^{-1}\bar{m}_0\Big\|_1 \le \frac{|\mu|}{\sqrt{\sigma^2 a+\mu^2}}\sqrt{z^\top G^{-1}z}\,\sum_{i=1}^n V_i.$$

对于值域 $\mathrm{Range}(B^\top)$ 中的分量，引理 4.14 还给出

$$\Big\|\bar{Q}(V)^{-1}\bar{m}_{\|}\Big\|_1 \le \sqrt{\sum_{i=1}^n V_i}\,\sqrt{\bar{m}_{\|}^\top \bar{Q}(V)^{-1}\bar{m}_{\|}}.$$

由于 $\bar{Q}(V)\succeq \rho B^\top B$，将 $\bar{Q}(V)^{-1}$ 限制到 $\mathcal{S}=$ 值域$(B^\top)$ 后满足 $\bar{Q}(V)^{-1}\preceq \frac{1}{\rho}(B^\top B)^\dagger$，因此

$$\bar{m}_{\|}^\top \bar{Q}(V)^{-1}\bar{m}_{\|} \le \frac{1}{\rho\lambda_+}\|\bar{m}_{\|}\|_2^2.$$

令 $S := \sum_{i=1}^n V_i$ 和 $c_{\|} := \frac{\|\bar{m}_{\|}\|_2}{\sigma\sqrt{\tilde{a}}\sqrt{\rho\lambda_+}}$，则

$$\frac{1}{\sqrt{\tilde{a}}}\Big\|\frac{\bar{Q}(V)^{-1}\bar{m}_{\|}}{\sigma}\Big\|_1 \le c_{\|}\sqrt{S}.$$

对于任意 $\varepsilon>0$，我们有 $c_{\|}\sqrt{S}\le \frac{\varepsilon}{2}S+\frac{c_{\|}^2}{2\varepsilon}$。

接下来，我们对高斯波动项进行界定。由于 $(\bar{Q}(V)^{-1})_{ii}\le V_i$，我们有

$$\mathbb{E}|Z_i| = \sqrt{\frac{2}{\pi}}\sqrt{(\bar{Q}(V)^{-1})_{ii}} \le \sqrt{\frac{2}{\pi}}\sqrt{V_i}.$$

Thus

$$\frac{1}{\sqrt{\tilde{a}}}\mathbb{E}\|Z\|_1 \le c_0\sum_{i=1}^n\sqrt{V_i}, \qquad c_0 := \sqrt{\frac{2}{\pi}}\frac{1}{\sqrt{\tilde{a}}}.$$

利用 $c_0\sqrt{V_i}\le \frac{\varepsilon}{2}V_i+\frac{c_0^2}{2\varepsilon}$ 并对 i 求和，可得

$$\frac{1}{\sqrt{\tilde{a}}}\mathbb{E}\|Z\|_1 \le \frac{\varepsilon}{2}\sum_{i=1}^n V_i + \frac{nc_0^2}{2\varepsilon}.$$

结合前述界，存在 $L_+<\infty$ 使得

$$\sum_{i=1}^n \mathbb{E}[\tilde{V}_i \mid V] \le \left(\frac{|\mu|}{\sqrt{\sigma^2 a+\mu^2}}\sqrt{z^\top G^{-1}z}+\varepsilon\right)\sum_{i=1}^n V_i + L_+.$$

根据假设 2.1，我们可以将 $\varepsilon > 0$ 选得足够小，使得

$$\gamma_+ := \frac{|\mu|}{\sqrt{\sigma^2 a + \mu^2}} \sqrt{z^\top G^{-1} z} + \varepsilon < 1.$$

最后，令 $\gamma := \max\{\gamma_-, \gamma_+\} < 1$ 且 $L := 1 + L_- + L_+$。那么，对所有 $V \in (0,\infty)^n$，$\mathbb{E}\Big[G_\delta(\tilde{V}) \mid V\Big] \le \gamma\, G_\delta(V) + L$，漂移不等式得证。 □

在建立漂移条件后，我们现在转而证明小化条件，从而完成对几何遍历性条件的验证。

引理 4.16（在 $\{G(V) \le d\}$上的小化）。固定 α，$\beta > 0$ 并定义

$$G(V) := \sum_{i=1}^{n} \big(V_i^{\alpha} + V_i^{-\beta}\big), \qquad C_d := \{V \in (0,\infty)^n : G(V) \le d\}.$$

假设对每个 $V \in (0,\infty)^n$，条件密度 $\pi(M \mid V, Y)$ 存在，对所有 $M \in \mathbb{R}^n$ 均严格为正，并且映射 $V \mapsto \pi(M \mid V, Y)$ 在 $(0,\infty)^n$ 上连续，其中 M 固定。（对于高斯更新 $M \mid (V, Y) \sim N(Q(V)^{-1}m, Q(V)^{-1})$，这些条件均成立。）令

$$k(V, \tilde{V}) := \int_{\mathbb{R}^n} \pi(\tilde{V} \mid M)\, \pi(M \mid V, Y)\, dM, \qquad V \in (0,\infty)^n,\ \tilde{V} \in (0,\infty)^n.$$

于是，对每个 $d > 0$，都存在 $\epsilon_d \in (0, 1]$ 以及定义在 $(0,\infty)^n$ 上的密度 $f_d(V)$，使得

$$k(V, \tilde{V}) \ \ge\ \epsilon_d\, f_d(\tilde{V}), \qquad \forall V \in C_d,\ \forall \tilde{V} \in (0,\infty)^n.$$

证明。首先，注意到 C_d 是紧集。事实上，对于任意 $V \in C_d$，都有 $V_i^{\alpha} \le G(V) \le d$，且对每个 i，$V_i^{-\beta} \le d$，这意味着 $d^{-1/\beta} \le V_i \le d^{1/\alpha}$。因此，$C_d$ 是 $[d^{-1/\beta}, d^{1/\alpha}]^n \subset (0,\infty)^n$ 的闭有界子集。

固定 $M \in \mathbb{R}^n$。由于映射 $V \mapsto \pi(M \mid V, Y)$ 在 $(0,\infty)^n$ 上连续，因此根据魏尔斯特拉斯定理，它在紧集 C_d 上的限制能取到最小值。定义

$$g_d(M) := \min_{V \in C_d} \pi(M \mid V, Y).$$

由于对于所有 $V \in (0,\infty)^n$，都有 $\pi(M \mid V, Y) > 0$，因此对所有 M，有 $g_d(M) > 0$。此外，任取 $V_0 \in C_d$，可得对所有 M，都有 $0 < g_d(M) \le \pi(M \mid V_0, Y)$，这表明 g_d 可积且满足

$$0 < \epsilon_d := \int_{\mathbb{R}^n} g_d(M)\, dM \le \int_{\mathbb{R}^n} \pi(M \mid V_0, Y)\, dM = 1.$$

令 $\tilde{g}_d(M) := g_d(M)/\epsilon_d$，它是 $\mathbb{R}^n$ 上的概率密度，并定义

$$f_d(\tilde{V}) := \int_{\mathbb{R}^n} \pi(\tilde{V} \mid M)\, \tilde{g}_d(M)\, dM.$$

根据托内利定理，并且由于 $\int_{(0,\infty)^n} \pi(\tilde{V} \mid M)\, d\tilde{V} = 1$，我们得到

$$\int_{(0,\infty)^n} f_d(\tilde{V})\, d\tilde{V} = \int_{\mathbb{R}^n} \tilde{g}_d(M)\, dM = 1,$$

这表明 f_d 是 $(0,\infty)^n$ 上的概率密度。

最后，对于任意 $V \in C_d$，我们有 $\pi(M \mid V, Y) \ge g_d(M)$ 对所有 M 均成立；因此，

$$k(V, \tilde{V}) = \int_{\mathbb{R}^n} \pi(\tilde{V} \mid M) \pi(M \mid V, Y)\, dM \ge \int_{\mathbb{R}^n} \pi(\tilde{V} \mid M)\, g_d(M)\, dM = \epsilon_d\, f_d(\tilde{V}),$$

这就是所需的小化。 □

建立这些初步结果后，我们现在可以证明命题 2.6。 .

命题 2.6 的证明。回顾一下，命题 2.6 根据漂移与小化条件，在三种状态下确立了几何遍历性。
情形 (I) 可由引理 4.7 和 4.16 通过定理 4.1 得出。情形 (II) 在假设 2.1 下可由引理 4.11 和 4.16 通过定理 4.1 得出。情形 (III) 在假设 2.1 下可由引理 4.15 和 4.16 通过定理 4.1 得出。

□

5. 模拟研究。

5.1. 概述与通用设置。 本节报告两项模拟研究（S1 和 S2），通过数值实验展示并对理论混合结果进行压力测试；这些理论混合结果已在第 4 节确立。两项研究均聚焦于由非中心化参数化诱导的 V 边际 Gibbs 链，因为关键理论机制（迹类性质以及漂移–小化）是针对 V 上的边缘转移算子提出的。我们使用标准的马尔可夫链蒙特卡罗效率诊断来比较混合行为。

模拟 S1 通过改变 GIG 参数 （p， a， b， μ），考察表 2 中不同参数区域下混合的变化情况。其目标是对比迹类情形（定理 2.1）与边界/重尾区域；在后者中，几何遍历性通过漂移–小集条件（命题 2.6）得以建立。模拟 S2 通过固定一个对边界敏感的分布情形并保持几何结构不变（即固定 A，从而固定 $B = AK$（ϕ）$^{-1}$），同时扫描漂移参数μ，考察零空间小性机制的实际影响，使零空间小性常数发生变化。混合情况通过对边界敏感和零空间敏感的汇总指标进行评估。

两项研究均采用式（5）中的非中心化参数化。令 $V \in$ （0， ∞）^{n} 表示混合变量，$M \in \mathbb{R}^n$ 表示潜在场。给定固定矩阵 A 和 K（ϕ）（其中 K（ϕ）可逆），定义 B : $= AK$（ϕ）$^{-1}$。模型为

$$
\begin{aligned}
Y \mid M &\sim N\big(X\beta + B(M - \mu h),\, \sigma_\epsilon^2 I\big),\\
M \mid V &\sim N\,bigl(\mu V,\, \mathrm{diag}(\sigma^2 V)\big),\\
V_i &\overset{\text{ind.}}{\sim} \mathrm{GIG}(p, a, b), \qquad i = 1, \ldots, n,
\end{aligned}
$$

其中我们使用密度为 $f(v; p, a, b) \propto v^{p-1}\exp\{-\frac{1}{2}(av + b/v)\}$的广义逆高斯分布参数化，适用于 $v > 0$。在此参数化下，非中心 Gibbs 采样器在高斯更新之间交替

$$
M \mid V, Y \sim N\big(\bar{Q}(V)^{-1}\bar{m},\, \sigma^2\bar{Q}(V)^{-1}\big), \qquad \bar{Q}(V) := \rho\, B^\top B + D_V^{-1}, \quad \rho := \sigma^2\sigma_\epsilon^{-2},
$$

以及相互独立的逐坐标 GIG 更新

$$
\tilde{V}_i \mid M_i \sim \mathrm{GIG}\Big(p - \tfrac{1}{2},\, a + \big(\tfrac{\mu}{\sigma}\big)^2,\, b + \big(\tfrac{M_i}{\sigma}\big)^2\Big), \qquad i = 1, \ldots, n,
$$

其中 $D_V = \mathrm{diag}(V_1, \ldots, V_n)$且 $\bar{m} := \rho\, B^\top\big(Y - X\beta + \mu B h\big) + \mu\mathbf{1}$。

除非另有说明，我们将问题规模固定为 $n = m = 300$，并设定 $h = \mathbf{1}$、 $X\beta = \mathbf{0}$、$\sigma = 1$、 $\sigma_\epsilon = 1$和 $Y = \mathbf{0}$。选择 $Y = \mathbf{0}$ 作为固定的基准观测值（而非从先验中抽样），以消除数据实现中的随机性，同时保留由 $\rho B^\top B$ 引入的依赖结构。对于 $K(\phi)$，我们采用参数为 $\phi = 0.5$ 的 AR(1) 型相关矩阵，并以数值方法计算 $K(\phi)^{-1}$。设计矩阵 A 在各项研究中分别设定（S1： $A = I_n$；S2： $A = I_n - uu^\top$）。

表 3模拟 *S1* 的代表性参数点。

点	状态	p	a	b	μ	键标签
A	TC-1	-0.5	1	1	1	IG 混合（$p=-\frac{1}{2}$）；正态逆高斯分布基准
B	TC-1	1.0	1	1	1	迹类内部（$a>0,\ b>0$）
C	TC-2	0.6	1	0	1	$b=0$ 边界；迹类（$p>\frac{1}{2}$） –
D	DM-III	0.3	1	0	1	$b=0$ 边界；非迹类（$0<p\leq\frac{1}{2}$） –
E	DM-I	-1.5	0	2	0	逆伽马分布的全条件分布（$a=\mu=0$）
F	DM-II	-1.5	0	2	1	漂移重尾（$a=0, p<0, \mu\neq 0$）

对于每条链及每个监测的标量统计量，我们计算积分自相关时间（IACT）和每秒有效样本量（ESS/秒），其中墙钟时间按每条链测量。对于保留长度为 N 的样本，$\mathrm{ESS}=N/\mathrm{IACT}$。为评估^条链之间的一致性，我们计算分裂-$R$，并按各项研究中的规定，报告所有监测统计量中的最大值。

5.2. 模拟研究（*S1*）：跨参数区域的混合比较。 本模拟研究从数值上说明 V边际 Gibbs 链的混合行为如何随表 2 所汇总的参数区域而变化。具体而言，我们将边缘马尔可夫算子 Λ 为迹类的区域，与迹类性不可得、且几何遍历性通过漂移–小集条件建立的边界/重尾区域进行对比。

我们采用第5.1节所述的通用模型、采样器与基准平台。在 S1 中，我们设定 $A=I_n$，从而 $B=K(\phi)^{-1}$，并仅在下文指定的代表性区域网格上改变 GIG 参数(p,a,b,μ)。我们评估覆盖表 2 中所有区域的六个代表性参数点。每个点指定(p,a,b,μ)，同时保持$(n,A,K(\phi)\ \sigma,\sigma_\epsilon,Y)$与第5.1节中相同：

这里，TC-1 和 TC-2 属于定理2.1中的迹类情形，而 DM-I–DM-III 则对应命题 2.6 中漂移–小集条件论证所涵盖的边界/重尾区域。

对于每个参数点（A–F），我们运行4 条过度离散链，进行 $T=50{,}000$ 次迭代，$T_{\mathrm{burn}}=5{,}000$，抽稀间隔为 $=1$。四个初始状态设为 $V^{(0)}=\mathbf{1}$, $V^{(0)}=0.1\,\mathbf{1}$, $V^{(0)}=10\,\mathbf{1}$，以及 $V^{(0)}\sim\mathrm{GIG}(1,1,1)$。我们不进行抽稀；所有报告的效率指标均基于保留的预热后抽样值。为将经验诊断与理论分析中使用的漂移函数联系起来，我们监测V-链的三个标量汇总量：

$$S_+(t):=\frac{1}{n}\sum_{i=1}^{n}V_i^{(t)},\qquad S_-(t):=\frac{1}{n}\sum_{i=1}^{n}\big(V_i^{(t)}\big)^{-q},\qquad S_{\log}(t):=\frac{1}{n}\sum_{i=1}^{n}\log V_i^{(t)},$$

与 $q=0.25$ 一起。统计量 S_+ 对右尾偏离较为敏感，而 S_- 和 $S_{\log}$对 $V\downarrow 0$ 附近的边界行为较为敏感，这在 $b=0$^情形中尤其重要。对于每个统计量，我们报告积分自相关时间、ESS/秒和 split-R，并用$\max\{R(S_+)\ R(S_-)\ R(S_{\log})\}$汇总一致性。根据理论结果，我们预期：（i）迹类情形（TC-1、TC-2）通常具有较高的抽样效率；（ii）具有 $b=0$ 和 $0<p\leq 1/2$（DM-III）的边界区域会出现最明显的减速，尤其是对于 S_- 和 $S_{\log}$。

表 4 汇报了六个代表性参数点（A–F）的效率诊断结果。^在所有状态和监测的汇总量中，分段-R 值都极为接近 1（其中

表 4*S1* 模拟汇总（积分自相关时间和 *ESS/秒*）。

状态	点	S_+		S_-		$S_{\log}$	
		IACT	ESS/秒	IACT	ESS/秒	IACT	ESS/秒
TC-1	A	1.31	90.91	1.33	90.02	1.34	89.07
TC-1	B	1.13	105.69	1.15	104.03	1.15	103.78
TC-2	C	1.48	80.71	1.80	67.89	2.28	52.53
DM-III	D	1.97	60.79	4.71	29.49	6.75	17.74
DM-I	E	1.02	119.83	1.17	104.37	1.17	104.33
DM-II	F	1.27	94.15	1.26	94.64	1.27	93.70

$\widehat{}$最大值 $R \approx 1.0001$），表明对于所考察的标量汇总量，各条链的表现一致。一个明显的定性规律是，$b = 0$ 边界区域会导致显著减速，尤其是对 $V \downarrow 0$敏感的汇总量。在迹类内部区域（TC-1，点 A–B）中，积分自相关时间接近 1，且在 S_+、S_-、$S_{\log}$方面的 ESS/秒较高。转到具有 $p > 1/2$ 的 $b = 0$（TC-2，点 C）时，混合程度有所下降，其中 $S_{\log}$最为明显。在具有 $0 < p \le 1/2$ 的 $b = 0$（DM-III，点 D）中，性能下降最为显著，此时 S_- 和 $S_{\log}$ 的积分自相关时间明显更长，且 ESS/秒最低。比较 DM-I 和 DM-II 可见，引入非零漂移参数 μ 会增加链中的依赖性：点 F（$\mu \neq 0$）的积分自相关时间系统性地大于点 E（$\mu = 0$），而 ESS/秒则更低，不过在本次实验中，两者的表现仍然良好。

这些实证发现与表 2 背后的机制一致。迹类状态通常具有更强的 $L_2(\pi)$ 谱性质，这往往会带来更高的抽样效率。当 $b = 0$ 和 p 较小（DM-III）时，目标分布在 $V \downarrow 0$ 附近表现出更明显的边界行为；相应地，对于负阶矩和$\log V$等靠近边界的函数量，链的混合速度较慢，这与漂移–小集分析中涉及 V负幂的李雅普诺夫函数相符。

5.3. 模拟研究（S2）：通过改变 μ进行零空间小性扫描。本模拟研究在受控条件下探究零空间小性常数如何影响 V边际 Gibbs 链的混合。我们固定几何结构（即固定 A，从而固定 $B = AK(\phi)^{-1}$），并扫描漂移参数 μ。改变 μ会同时改变组合量 $\tilde{a}(\mu) = a + \big(\mu/\sigma\big)^2$以及高斯更新中的驱动项 $\bar{m}(\mu)$，从而构成一维压力测试：在全局耦合由 $B^\top B$ 决定且保持不变的情况下，零空间小性常数随之变化。我们采用第5.1节所述的常用模型、非中心 Gibbs 采样器和基准平台。在 S2 中，我们通过设定 $A := I_n - uu^\top$ 且 $u := \mathbf{1}/\|\mathbf{1}\|_2$，将设计矩阵固定为具有一维零空间的矩阵，使得 $\mathrm{Null}(A) = \mathrm{span}\{u\}$，并且 $B := AK(\phi)^{-1}$在整个实验中保持不变（其中 $K(\phi)$是 $\phi = 0.5$处的 AR(1) 型相关矩阵）。我们采用基准 $Y = \mathbf{0}$ 和 $X\beta = \mathbf{0}$，因此对 μ 的依赖仅通过漂移项体现。

我们在固定网格 $\mu \in \{\mu_1, \dots, \mu_L\}$ 上扫描漂移参数，并针对每个 μ 计算与固定算子 B 相关的零空间小性常数。设 u_0 为张成 $\mathrm{Null}(B)$ 的单位向量。对于 $B = AK(\phi)^{-1}$，其中 $A = I - uu^\top$，有 $\mathrm{Null}(B) = \mathrm{span}\{K(\phi)u\}$；因此我们取 $u_0 := K(\phi)u/\|K(\phi)u\|_2$。回顾 $\tilde{a}(\mu) := a + \big(\mu/\sigma\big)^2$、$\bar{m}(\mu) := \rho B^\top\big(Y - X\beta + \mu Bh\big) + \mu\mathbf{1}$ 以及 $\rho := \sigma^2\sigma_\epsilon^{-2}$。我们定义

$$\gamma_{\mathrm{ns}}(\mu) := \frac{\|P_{\mathrm{Null}(B)}\bar{m}(\mu)\|_2}{\sigma\sqrt{\tilde{a}(\mu)}} = \frac{\big|\langle u_0, \bar{m}(\mu)\rangle\big|}{\sigma\sqrt{\tilde{a}(\mu)}}, \tag{36}$$

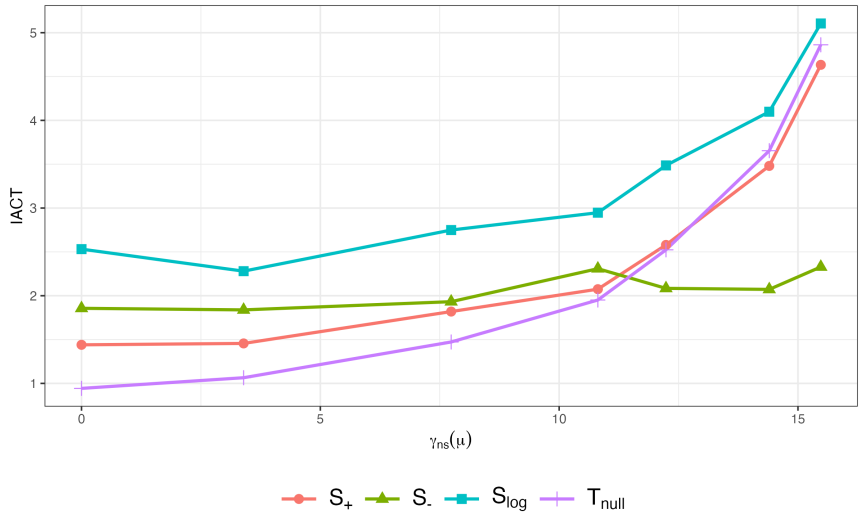

图 1：模拟 S2：在改变 μ、同时固定 A 的情况下，沿零空间小性扫描观察到的混合效率。图中报告了四种监测统计量的积分自相关时间：S_+（右尾）、S_- 和 $S_{\log}$（对边界敏感），以及零方向统计量 T_{null}。

其中，第二个恒等式利用了 $\mathrm{Null}(B)$ 是一维的这一事实。我们固定一个边界敏感区域，仅改变 μ。在实现中，我们聚焦于 $b=0$ 边界状态（DM- III），取 $(p,a,b)=(0.5,1,0)$ 和 $h=\mathbf{1}$。我们监测 S1 研究中引入的三个标量汇总量 S_+、S_- 和 $S_{\log}$，以及零空间方向统计量

$$T_{\text{null}}(t) := \big\langle K(\phi)^{-1}u_0,\ M^{(t)} - \mu h \big\rangle. \tag{37}$$

对于每个统计量和每条链，我们都计算积分自相关时间及 split-$\widehat{R}$。

对于扫描网格中的每个 μ，我们运行 4 条过度离散链，进行 $T=50{,}000$ 次迭代，$T_{\text{burn}}=10{,}000$，并以 $=1$ 抽稀；初始值设为 $V(0)=\mathbf{1}$、$V(0)=0.1\,\mathbf{1}$、$V(0)=10\,\mathbf{1}$，以及$V(0)\sim$广义逆高斯分布(**1**,1,1)。所有效率指标均根据保留的预热后抽样值计算，并记录每条链的墙钟时间。

在所有扫描中，所有监测的 $\widehat{R}$ 值均保持为 1，表明过度离散链之间的行为一致。增大 $\gamma_{\text{ns}}(\mu)$ 伴随着采样效率的系统性下降，这表现为多个汇总统计量的积分自相关时间不断增加，尽管通过 $B^\top B$ 保持了固定的全局耦合。在此次扫描中，零空间方向统计量 $T_{\text{null}}(t)$ 和对数汇总统计量 $S_{\log}(t)$ 对 $\gamma_{\text{ns}}(\mu)$ 表现出最强的敏感性，其积分自相关时间在整个扫描范围内显著增加。负矩汇总统计量 $S_-(t)$ 通常变化较为温和，这表明此处占主导地位的减速机制与沿公式 (37) 所编码的零空间敏感方向进行探索有关，而不仅仅是 V 的近边界偏离。

观测到的性能恶化与零小性条件背后的基于漂移的机制一致：扫描 μ 会同时改变 $\bar{m}(\mu)$ 和 $\tilde{a}(\mu)$，从而改变 $\bar{m}(\mu)$ 在 $\mathrm{Null}(B)$ 上的归一化投影，该投影由 $\gamma_{\text{ns}}(\mu)$ 捕获。这些结果为以下数值事实提供了证据：该常数取值越大，零空间敏感泛函的混合可能越慢，从而补充了第 4 节所建立的条件。

6. 讨论。 本文为线性潜变量非高斯模型（LLnGMs）的推断建立了一套连贯的理论与计算基础：通过方差混合增广引入非高斯性，同时保留条件高斯

结构。本文的主要贡献是提出了一组可验证的条件，在这些条件下，马尔可夫链具有强稳定性（几何遍历性），从而支持蒙特卡罗估计和基于随机梯度下降的似然优化。

本文的核心观点是，可以通过两种互补机制证明几何遍历性，而每种机制都揭示了链的不同方面。定理2.1中的迹类结果为混合变量上的边缘链快速混合提供了算子理论层面的判据。当相应的积分算子属于迹类时，便可获得良好的谱正则性；尤其是，由吉布斯更新诱导的马尔可夫链具有几何遍历性。从概念上说，这一路径是全局性的：它依赖于转移算子的紧性和可和性，并且在条件密度具有足够正则化的参数范围内，只需要相对温和的分布层面假设。漂移与小化方法则补充了迹类论证，针对定理2.1未涵盖的参数范围。在这些情况下，通过构造李雅普诺夫函数，并在状态空间的适当区域上建立小集条件，即可推出几何遍历性。这一路径具有局部性和模块化特点：它明确说明模型的不同组成部分如何影响稳定性，并且可以加以调整，以处理边界参数范围；在这些范围内，混合分布在0 附近可能更加奇异，或具有更重的尾部。在实际应用中，漂移论证还能定量揭示，为确保压缩算子性质必须控制哪些矩。综合这两种方法，可形成清晰的证明策略：在条件允许时采用迹类结果，以简洁地得出结论；否则，使用漂移–小集条件处理剩余的边界情形。这种分工既避免了在性质良好的参数范围内引入过于技术性的假设，也保留了对实际相关参数化的覆盖。

遍历性分析中的某些条件，通过潜变量算子与观测算子的几何结构来理解最为自然。尤其是，零空间的相互作用（例如 零空间(B) 及相关投影算子）决定了潜在场的某些方向是否受到数据的弱约束，进而可能导致混合更新近乎退化。漂移分析中出现的零空间小性条件，将这样一项要求形式化：观测机制应向那些原本会受到先验算子弱约束的潜在方向注入足够的信息。从这个意义上说，这些假设并非纯粹的技术条件，而是体现了一项具体的可识别性原理。此外，界中涉及的基于投影的常数（例如 $\|P_{\mathrm{Null}(B)}\mathbf{1}\|_2$ 及其可计算的替代量）为诊断特定离散化中的问题状态提供了一种方法。

更广泛地说，本文提出的证明策略并不局限于采用广义逆高斯分布（GIG）混合分布这一特定选择。只要混合变量在边界附近和尾部的渐近行为能够得到适当控制，该策略便可应用于更广泛的正态均值—方差混合模型。从技术角度看，遍历性论证仅依赖于混合变量条件密度存在上、下包络界，且这些界与 GIG 型核的线性组合具有可比性。具体而言，漂移与小化的构造要求明确控制 $V \to 0$ 和 $V \to \infty$处混合密度的主导阶渐近行为，并保证有限个逆矩和正矩的可积性。因此，迹类与基于漂移的论证可视为适用于具有重尾或偏斜边缘分布的条件高斯潜变量模型中 Gibbs 采样器的结构性结果，而非针对特定分布的证明。这一观察表明，只需验证混合分布的一小组渐近条件，无须从头重新进行完整的算子或漂移分析，便可能将本文建立的理论框架扩展到其他混合构造。从这个意义上说，本文为范围广泛得多的潜变量非高斯模型提供了一套可重复使用的稳定性分析蓝图。

几何遍历性不仅是关于收敛的定性陈述；它还支撑着似然优化中所用蒙特卡罗估计量的有效性与可靠性。

具体而言，借助费舍尔恒等式与拉奥–布莱克韦尔化得到的随机梯度估计量，要求在不变分布下具有可积性，并且Gibbs链具有足够快的混合。本文的结果在所关注的参数区域内提供了这些稳定性保证，从而为使用遍历平均进行梯度估计提供了依据，并支持建立在该采样器之上的随机优化算法的收敛性论断。从实现角度看，这些结果促使我们采用能够保持条件高斯结构、同时改善混合的参数化与更新分块。尽管中心化与非中心化Gibbs核以同一后验分布为目标，它们的混合行为可能有显著差异。当数据对潜在场的识别较弱时，非中心化参数化可以降低后验耦合；而当似然信息充分、潜在场受到严格约束时，中心化参数化可能更为可取。本文的分析强调，推断所需的稳定性性质对两种表述均可建立，并阐明哪些分布情形更自然地由算子理论论证处理，哪些则由漂移构造处理。这一视角给出了实用准则：采用经验上能降低自相关的参数化，同时确保所得链仍处于理论所覆盖的状态之内。

理论结果针对通过方差混合增广保持条件高斯性的 LLnGM。打破共轭性的扩展（例如非线性观测模型，或没有辅助变量表示的非高斯似然）可能需要不同的马尔可夫链构造和新的稳定性论证。此外，尽管漂移–小化方法适用范围广，但它可能给出较保守的常数，也未必能直接导出关于蒙特卡洛误差的尖锐非渐近界。

致谢。 本出版物基于沙特阿卜杜拉国王科技大学（KAUST）在项目编号 ORFS-CRG11-2022-5015 下资助的研究工作。

附录A：第二类修正贝塞尔函数

我们总结第二类修正贝塞尔函数 $K_\nu(x)$ 的若干有用性质。以下渐近性质可见于 [1], 第 375 页：

$$K_\nu(x) = K_{-\nu}(x) \tag{38}$$

$$K_\nu(x) \sim \sqrt{\frac{\pi}{2x}} e^{-x} \text{ as } x \to \infty \tag{39}$$

$$K_\nu(x) \sim \frac{\Gamma(\nu)}{2} \left(\frac{x}{2}\right)^{-\nu} \text{ for } \nu > 0 \text{ as } x \to 0 \tag{40}$$

$$K_0(x) \sim -\log(x) \text{ as } x \to 0. \tag{41}$$

此外，我们的分析还依赖于 [21] 中建立的若干比值不等式。

命题 A.1（修正贝塞尔函数不等式 [21, 命题 A2]）。令 $\nu_1,\ \nu_2 > 0$。则对于任意 $\epsilon^* > 0$，存在 $\epsilon > 0$，使得

$$\frac{K_{\nu_1}(x)}{K_{\nu_2}(x)} \leq (1+\epsilon^*) \frac{\Gamma(\nu_1) 2^{\nu_1-\nu_2}}{\Gamma(\nu_2) x^{\nu_1-\nu_2}}$$

对所有 $0 < x < \epsilon$，其中 ϵ 取决于 ϵ^*、ν_1 和 ν_2。

命题 A.2（修正贝塞尔函数不等式 [21, 命题 A3]）。设 $\nu > 0$。则存在 $\epsilon' > 0$ 使得

$$\frac{K_\nu(x)}{K_0(x)} \le \frac{2^{\frac{1}{4}}\sqrt{\pi}}{2\Gamma(\frac{1}{4})x^\nu}$$

对于所有 $0 < x < \epsilon'$，其中 ϵ' 取决于 ν。

命题 A.3（修正贝塞尔函数不等式 [21, 命题 A4]）。设 $\nu \ge 0$、$\delta \in (0, 1)$ 且 $x > 0$。则

$$\frac{K_{\nu+\delta}(x)}{K_\nu(x)} \le \frac{(2\nu+1)^{\frac{\delta}{2}}}{x^{\frac{\delta}{2}}} + 1.$$

APPENDIX B: PROOFS

B.1. 第 2 节的证明。

命题 2.1 的证明。两种参数化的推导都依赖于高斯似然相对于高斯过程先验的共轭性，以及广义逆高斯分布对于高斯方差混合模型中的方差分量具有共轭性。从 $W \mid V$，Y的中心化参数化开始，观测值与潜在过程的联合分布由 $Y \mid W$，$V \sim N$（$X\beta + AW$，$\sigma_\epsilon^2 I$）和 $KW \mid V \sim N$（μ（$V - h$） $\sigma^2 D_V$）给出。条件密度 p（$W \mid V$，Y）与相应核的乘积成正比：

$$p(W \mid V, Y) \propto \exp\Bigg\{ -\frac{1}{2\sigma_\epsilon^2}\|Y - X\beta - AW\|^2 - \frac{1}{2\sigma^2}(KW - \mu(V-h))^\top D_V^{-1}(KW - \mu(V-h)) \Bigg\}.$$

展开二次型并收集涉及 W 的项，可得到高斯核$\exp\{-\frac{1}{2}(W^\top Q W - 2W^\top r)\}$，其中精度矩阵 Q和势向量 r定义为

$$Q = \sigma^{-2}K^\top D_V^{-1}K + \sigma_\epsilon^{-2}A^\top A, \quad r = \sigma^{-2}K^\top D_V^{-1}\mu(V-h) + \sigma_\epsilon^{-2}A^\top(Y - X\beta).$$

配方后得到条件分布 $W \mid V, Y \sim N(Q^{-1}r, Q^{-1})$，从而证实了（6）。

对于非中心化参数化 $M \mid V, Y$，我们考虑变换 $W = K^{-1}(M - \mu h)$。定义 $B = AK^{-1}$后，Y的似然可简化为 $N(X\beta - \mu Bh + BM, \sigma_\epsilon^2 I)$，而 M的先验为 $N(\mu V, \sigma^2 D_V)$。因此，条件密度 $p(M \mid V, Y)$ 为：

$$p(M \mid V, Y) \propto \exp\Bigg\{ -\frac{1}{2\sigma_\epsilon^2}\|Y - X\beta + \mu Bh - BM\|^2 - \frac{1}{2\sigma^2}(M - \mu V)^\top D_V^{-1}(M - \mu V) \Bigg\}.$$

再次整理 M 中的各项，可得到一个高斯核，其形式为

$$Q = \sigma_\epsilon^{-2}B^\top B + \sigma^{-2}D_V^{-1}, \quad r = \sigma_\epsilon^{-2}B^\top(Y - X\beta + \mu Bh) + \sigma^{-2}D_V^{-1}\mu V.$$

利用恒等式 $D_V^{-1}V = \mathbf{1}$，线性项可简化为 $r = \sigma_\epsilon^{-2}B^\top$（$Y - X\beta + \mu Bh$）+（$\mu/\sigma^2$）$\mathbf{1}$，这与（7）一致。

关于方差分量 V 的更新，注意在这两种情况下，一旦给定潜在过程， Y 与 V 条件独立。对于中心化情形 $V \mid W$， $(V_i)_i$ 的独立性使我们能够将密度分解为 $p(V \mid W) \propto \prod_i p(V_i)\, p(z_i \mid V_i)$ ，其中 $z = KW$。对于每个 i，核为

$$p(V_i \mid W) \propto V_i^{p-1} e^{-\frac{1}{2}(aV_i + b/V_i)} \cdot V_i^{-1/2} \exp\left\{-\frac{(z_i + \mu h_i - \mu V_i)^2}{2\sigma^2 V_i}\right\}.$$

展开指数项，并仅保留依赖于 V_i 的项，得到

$$p(V_i \mid W) \propto V_i^{p-1/2-1} \exp\left\{-\frac{1}{2}\left[\left(a + \frac{\mu^2}{\sigma^2}\right) V_i + \left(b + \frac{(z_i + \mu h_i)^2}{\sigma^2}\right) \frac{1}{V_i}\right]\right\},$$

这就是广义逆高斯分布$(p - 1/2, a + \mu^2/\sigma^2, b + (z_i + \mu h_i)^2/\sigma^2)$ 分布的核。

最后，对于非中心化情形 $V \mid M$，将先验 $M_i \mid V_i \sim N(\mu V_i, \sigma^2 V_i)$ 展开，并结合 V_i 上的广义逆高斯先验，可得

$$p(V_i \mid M) \propto V_i^{p-1/2-1} \exp\left\{-\frac{1}{2}\left[\left(a + \frac{\mu^2}{\sigma^2}\right) V_i + \left(b + \frac{M_i^2}{\sigma^2}\right) \frac{1}{V_i}\right]\right\},$$

这证实了式 **(9)** 中的表达式，并完成了证明。 □

B.2. 第3节的证明。

引理 3.1 的证明。设 $Z = KW$ 和 $S_i = V_i^{-1} h_i (Z_i + \mu h_i)$。为简化表述，我们固定索引 i，并在可能时将其省略，分别用 v, h, z, w 表示 V_i, h_i, Z_i, W_i。在整个证明过程中，我们用小写字母表示积分变量。根据贝叶斯定理， $|S_i|$ 的后验期望为

$$\mathbb{E}[|S_i| \mid y] = \frac{1}{\pi(y)} \int \frac{|h|\,|z + \mu h|}{v} p(y \mid w) p(z, v)\, dz\, dv.$$

我们首先证明，当 $\alpha > \frac{1}{2}$ 时，积分有限。由于高斯似然在全局上有界， $p(y \mid w) \le L_{\max} := (2\pi\sigma_\epsilon^2)^{-n/2}$, 我们得到 $\mathbb{E}[|S_i| \mid y] \le \frac{L_{\max}}{\pi(y)} \mathbb{E}\left[\frac{|h|\,|z+\mu h|}{V}\right]$ · 定义 $T := z_i + \mu h$，我们有 $T \mid v \sim \mathcal{N}(\mu v, \sigma^2 v)$，因此 $\mathbb{E}(|T| \mid v) \le c_1 \sqrt{v} + |\mu| v$ 其中 $c_1 := \sigma\sqrt{2/\pi}$，于是

$$\mathbb{E}\left[\frac{|h|\,|T|}{v}\right] \le |h| \int_0^\infty (c_1 v^{-1/2} + |\mu|)\, p_V(v)\, dv, \quad p_V(v) \propto v^{\alpha-1} e^{-\beta v}.$$

当 0 时，被积函数的行为类似于 $v^{-1/2} v^{\alpha-1} = v^{\alpha-3/2}$，因此该积分有限当且仅当 $\alpha > \frac{1}{2}$。

现在我们证明，对于 $\alpha \le \frac{1}{2}$，期望值是发散的。令 $\tilde{z} := z_{-i}$。我们希望找到一个集合，使得可以从下方界定积分中的各项 $\mathbb{E}(|S_i| \mid y)$。令 $z^\star := -\mu h \mathbf{1}$，并选择 $\eta > 0$。定义立方体

$$Q := \{z \in \mathbf{R}^n : \|z - z^\star\|_\infty \le \eta\}, \qquad Q_{-i} := \{\tilde{z} \in \mathbf{R}^{n-1} : \|\tilde{z} - \tilde{z}^\star\|_\infty \le \eta\}.$$

令 $B := K^{-1} Q$。由于 $w \mapsto p(y \mid w)$ 连续且严格为正，因此存在 $L_{\min} :=$ 下确界 $_{w \in B}\, p(y \mid w) > 0$. 因此，对所有 $z \in Q$，我们都有 $p(y \mid K^{-1} z) \ge L_{\min}$，从而

$$\begin{aligned}\mathbb{E}[|S_i| \mid y] &= \frac{1}{\pi(y)} \int \frac{|h|\,|z_i + \mu h|}{v}\, p(y \mid K^{-1} z)\, p(z, v)\, dz\, dv \\ &\ge \frac{L_{\min}}{\pi(y)} \int_Q \int \frac{|h|\,|z_i + \mu h|}{v}\, p(z, v)\, dz\, dv.\end{aligned}$$

固定 $b>0$，并定义 $I(v) := [b\sigma\sqrt{v},\ 2b\sigma\sqrt{v}\,]$. 选取 $\varepsilon = b/4$，并挑选 $\delta>0$ 使得 $2b\sigma\sqrt{\delta}\le\eta$ 且 $\frac{|\mu|}{\sigma}\sqrt{\delta}\le\varepsilon$（稍后使用）。那么对 $v\in(0,\delta)$，$t\in I(v)$ 且 $\tilde z\in Q_{-i}$，我们有 $(z_i^\star + t, \tilde z)\in Q$。由此得到

$$\begin{aligned}\int_Q\int \frac{|h|\,|z_i+\mu h|}{v}\,p(z,v)\,dz\,dv &\\ &\ge \int_0^\delta \int_{Q_{-i}}\int_{t\in I(v)} \frac{|h|\,t}{v}\,p_{Z_i|V_i}(z_i^\star+t\mid v)\,p_V(v)\prod_{j\ne i}p_{Z_j}(z_j)\,dt\,d\tilde z\,dv\\ &= c_{-i}\int_0^\delta\left[\int_{t\in I(v)}\frac{|h|\,t}{v}\,p_{Z_i|V_i}(z_i^\star+t\mid v)\,dt\right]p_V(v)\,dv,\end{aligned}$$

其中 $c_{-i} := \int_{Q_{-i}}\prod_{j\ne i}p_{Z_j}(z_j)\,d\tilde z\ >0.$

对于 $t\in I(v)$，我们有 $t/v\ge b\sigma v^{-1/2}$。因此，对于所有足够小的 v,

$$\begin{aligned}&\int_0^\delta\left[\int_{t\in I(v)}\frac{|h|\,t}{v}\,p_{Z_i|V_i}(z_i^\star+t\mid v)\,dt\right]p_V(v)\,dv\\ &\quad\ge\ \int_0^\delta b\sigma|h|\,v^{-1/2}\left[\int_{t\in I(v)}p_{Z_i|V_i}(z_i^\star+t\mid v)\,dt\right]p_V(v)\,dv.\end{aligned}$$

令 $s(v) := \frac{\mu}{\sigma}\sqrt{v}$，由于 $\frac{|\mu|}{\sigma}\sqrt{\delta}\le\varepsilon$，我们有 $|s(v)|\le\varepsilon$，对所有 $v\in(0,\delta)$ 都成立。令 $f(s) := \Phi(2b-s)\ -\Phi(b-s)$，并定义 $p_0 := $ 最小值$_{|s|\le\varepsilon}\,f(s)\ >0$，由此得到

$$\begin{aligned}\int_{t\in I(v)}p_{Z_i|V_i}(z_i^\star+t\mid v)\,dt &= \int_{b\sigma\sqrt{v}}^{2b\sigma\sqrt{v}}\frac{1}{\sqrt{2\pi\sigma^2 v}}\exp\!\left(-\frac{(t-\mu v)^2}{2\sigma^2 v}\right)dt\\ &= f(s(v))\ \ge\ p_0, \forall v\in(0,\delta).\end{aligned}$$

因此

$$\mathbb{E}[|S_i|\mid y]\ \ge\ c\int_0^\delta v^{-1/2}\,p_V(v)\,dv\ \asymp\ \int_0^\delta v^{-1/2}v^{\alpha-1}\,dv=\int_0^\delta v^{\alpha-3/2}\,dv=\infty\qquad(\alpha\le\tfrac12),$$

如所声称。 □

B.3. 第 4 节 4.1 的证明。

引理 4.1 的证明。首先，我们建立一个有用的矩阵恒等式。由于 A 和 B 是正定的，因此它们可逆；它们的和 $A+B$ 以及它们的逆矩阵之和$A^{-1}+B^{-1}$ 也都是正定的（因而可逆）。我们可以写出

$$A^{-1}+B^{-1}=A^{-1}(A+B)B^{-1}.$$

对等式两边取逆，得到恒等式

$$(A^{-1}+B^{-1})^{-1}=B(A+B)^{-1}A, \tag{42}$$

由对称性可知，它也等于 $A(A+B)^{-1}B$。利用表示式 $I=(A+B)(A+B)^{-1}$，我们可以将涉及 A 的项改写为

$$\begin{aligned}A-(A^{-1}+B^{-1})^{-1} &= A-A(A+B)^{-1}B\\ &= A(A+B)^{-1}(A+B)-A(A+B)^{-1}B=A(A+B)^{-1}A.\end{aligned}$$

类似地，对于涉及 B 的项，我们有 $B-(A^{-1}+B^{-1})^{-1}=B(A+B)^{-1}B$。现在，让我们展开引理左侧的表达式。令 $H=(A^{-1}+B^{-1})^{-1}$。我们有

$$\begin{aligned}u^\top Au+v^\top Bv-(u+v)^\top H(u+v)&=u^\top Au+v^\top Bv-(u^\top Hu+v^\top Hv+2u^\top Hv)\\&=u^\top(A-H)u+v^\top(B-H)v-2u^\top Hv=(I).\end{aligned}$$

将上文推导出的恒等式代入该表达式：

$$\begin{aligned}(I)&=u^\top\big(A(A+B)^{-1}A\big)u+v^\top\big(B(A+B)^{-1}B\big)v-2u^\top\big(A(A+B)^{-1}B\big)v\\&=(Au)^\top(A+B)^{-1}(Au)+(Bv)^\top(A+B)^{-1}(Bv)-2(Au)^\top(A+B)^{-1}(Bv)\\&=(Au-Bv)^\top(A+B)^{-1}(Au-Bv).\end{aligned}$$

由此证明了该等式。由于 $A+B$ 是正定的，最后一行的二次型非负，从而得到所需的不等式。 □

引理4.2的证明。定义

$$f(x):=K_p(x)\,x^{|p|}(1+x)^{\frac12-|p|}e^x,\qquad x>0.$$

由于对 $x>0$有 $K_p(x)>0$，故对所有 $x>0$有 $f(x)>0$。此外，f 在$(0,\infty)$上连续，因为 K_p在$(0,\infty)$上连续。利用 K_p的标准渐近性质（见附录A）：

- 当 $x\to\infty$时，$K_p(x)\sim\sqrt{\frac{\pi}{2x}}e^{-x}$，因此

$$f(x)\to\sqrt{\frac{\pi}{2}}\cdot\lim_{x\to\infty}\frac{x^{|p|}(1+x)^{\frac12-|p|}}{\sqrt{x}}=\sqrt{\frac{\pi}{2}}>0.$$

当 $x\to0^+$时，若 $p\neq0$，则 $K_p(x)\sim2^{|p|-1}\Gamma(|p|)\,x^{-|p|}$，因此 $f(x)\to2^{|p|-1}\Gamma(|p|)>0$；而若 $p=0$，则 $K_0(x)\to\infty$，所以 $f(x)\to\infty$。

因此，存在 $\varepsilon\in(0,1)$ 和 $M>1$，使得

$$f(x)\geq c_0>0\quad\text{for all } x\in(0,\varepsilon]\cup[M,\infty).$$

在剩余的紧区间 $[\varepsilon,M]$ 上，由连续性可知，f 能够取得其最小值：

$$m:=\min_{x\in[\varepsilon,M]}f(x)>0,$$

因为 $f>0$ 处处成立。因此

$$\inf_{x>0}f(x)\;\geq\;\min\{c_0,m\}\;>\;0.$$

令 $C:=$最小值$\{c_0,m\}$（或者任意一个 $0<C<$下确界$_{x>0}f(x)$，如果你坚持使用严格不等式）。那么对于所有 $x>0$，$f(x)\geq C$，这等价于 $K_p(x)\;\geq\;C\,x^{-|p|}(1+x)^{-\frac12+|p|}e^{-x}$。□

引理4.3的证明。设 φ表示分布 $N(A^{-1}b,A^{-1})$的密度。则该结论等价于证明

$$\int\prod_{i=1}^n(\gamma+x_i^2)^p\,\varphi(x)\,dx\leq C\prod_{i=1}^n\Big(\gamma^p+[A^{-1}]_{ii}^p\big(1+\sum_{j=1}^n[A^{-1}]_{jj}^p\big)\Big),\tag{43}$$

其中 $C=C(p,n,\|b\|)$。令 $k>np$为一个整数。根据赫尔德不等式，

$$\int\prod_{i=1}^n(\gamma+x_i^2)^p\varphi(x)\,dx\leq\prod_{i=1}^n\left(\int(\gamma+x_i^2)^{np}\varphi(x)\,dx\right)^{1/n}.$$

由于 $k > np$， L^r 范数的单调性可得

$$\int (\gamma + x_i^2)^{np} \varphi(x)\, dx \le \left(\int (\gamma + x_i^2)^k \varphi(x)\, dx \right)^{np/k}.$$

结合这两个不等式，

$$\int \prod_{i=1}^n (\gamma + x_i^2)^p \varphi(x)\, dx \le \prod_{i=1}^n \left(\int (\gamma + x_i^2)^k \varphi(x)\, dx \right)^{p/k}.$$

令 $m = A^{-1}b$。利用不等式 $x_i^2 \le 2m_i^2 + 2(x_i - m_i)^2$，我们得到

$$(\gamma + x_i^2)^k \le (\gamma + 2m_i^2 + 2(x_i - m_i)^2)^k \le 3^{k-1}\big(\gamma^k + 2^k m_i^{2k} + 2^k (x_i - m_i)^{2k}\big).$$

对 φ 取期望值并将常数吸收到 C中，

$$\int (\gamma + x_i^2)^k \varphi(x)\, dx \le C\big(\gamma^k + |m_i|^{2k} + [A^{-1}]_{ii}^k\big),$$

由于 $\mathbb{E}_\varphi[(x_i - m_i)^{2k}] = C_k [A^{-1}]_{ii}^k$。将两边提升至 p/k 次方，并利用$p/k \le 1$，我们得到

$$\left(\int (\gamma + x_i^2)^k \varphi(x)\, dx \right)^{p/k} \le C\big(\gamma^p + |m_i|^{2p} + [A^{-1}]_{ii}^p\big).$$

最后，由于 $m = A^{-1}b$ 且 b 是固定的，我们有

$$|m_i| \le \|b\| \sum_{j=1}^n |[A^{-1}]_{ij}| \le \|b\| [A^{-1}]_{ii}^{1/2} \sum_{j=1}^n [A^{-1}]_{jj}^{1/2},$$

这意味着

$$|m_i|^{2p} \le C(\|b\|)\, [A^{-1}]_{ii}^p \sum_{j=1}^n [A^{-1}]_{jj}^p.$$

将此上界代入（43），即可得到所需结果。 □

B.4. 第 4.2 节的证明。

引理 4.4 的证明。在此设定下，边际数据增广算法对应于算子 Λ，其联合密度为 $\pi(V,\ M)$。Λ为迹类的一个必要条件是该积分有限：

$$T := \int \int \pi(V_i \mid M_i) \pi(M_i \mid V_i)\, dV_i\, dM_i.$$

我们证明该积分发散。

在假设 $a = \mu = 0$下，我们有 $p < 0$。令 $\alpha := -p > 0$ 且 $\beta := b/2 > 0$。V_i的先验为 $V_i \sim$ 逆伽马分布(α, β)。当 $B = 0$时，条件分布简化为：

$$M_i \mid V_i \sim N(0, V_i),$$

$$V_i \mid M_i \sim \text{Inv-Gamma}\left(\alpha + \frac{1}{2}, \beta + \frac{M_i^2}{2} \right).$$

相应的密度函数为

$$\pi(M_i \mid V_i) = \frac{1}{\sqrt{2\pi V_i}} \exp\left(-\frac{M_i^2}{2V_i}\right),$$

$$\pi(V_i \mid M_i) = \frac{(\beta + M_i^2/2)^{\alpha+1/2}}{\Gamma(\alpha+1/2)} V_i^{-\alpha-3/2} \exp\left(-\frac{\beta + M_i^2/2}{V_i}\right).$$

将这些代入 T 的表达式，并应用富比尼定理，先对 V_i 积分:

$$T = \int_{-\infty}^{\infty} \frac{(\beta + M_i^2/2)^{\alpha+1/2}}{\sqrt{2\pi}\,\Gamma(\alpha+1/2)} \left[\int_0^{\infty} V_i^{-\alpha-2} \exp\left(-\frac{\beta + M_i^2/2}{V_i}\right) dV_i\right] dM_i.$$

内部积分与未归一化的逆伽马密度有关。利用恒等式 $\int_0^{\infty} x^{-(k+1)} e^{-C/x} dx = \Gamma(k) C^{-k}$，其中 $k = \alpha + 1$ 且 $C = \beta + M_i^2/2$，可得:

$$\int_0^{\infty} V_i^{-(\alpha+1)-1} \exp\left(-\frac{\beta + M_i^2/2}{V_i}\right) dV_i = \Gamma(\alpha+1) \left(\beta + \frac{M_i^2}{2}\right)^{-(\alpha+1)}.$$

将其代回 T 的表达式:

$$T = \frac{\Gamma(\alpha+1)}{\sqrt{2\pi}\,\Gamma(\alpha+1/2)} \int_{-\infty}^{\infty} \left(\beta + \frac{M_i^2}{2}\right)^{\alpha+1/2} \left(\beta + \frac{M_i^2}{2}\right)^{-(\alpha+1)} dM_i$$

$$= \frac{\Gamma(\alpha+1)}{\sqrt{2\pi}\,\Gamma(\alpha+1/2)} \int_{-\infty}^{\infty} \left(\beta + \frac{M_i^2}{2}\right)^{-1/2} dM_i.$$

当 $|M_i| \to \infty$ 时，被积函数的行为类似于 $(\frac{M_i^2}{2})^{-1/2} \propto \frac{1}{|M_i|}$。由于 $\int_1^{\infty} \frac{1}{x} dx$ 按对数速度发散，积分 T 为无穷大。因此，该算子不是迹类。 □

以下两个引理给出证明引理 4.5 所需的辅助结果。为了陈述这些结果，我们需要先定义几个概念。设 $A \in \mathbb{R}^{n\times n}$。行的指标即其顺序，也就是说，第 $[a_{i,j}]_{j\in\{1,\dots,n\}}$ 行的指标为 i；列的指标类似定义。设 $S_r \subseteq \{1,2,\dots,n\}^r$ 为由 $\{1,\dots,n\}$ 中的元素构成、长度为 r 的严格递增序列集合。对于 $r \in \{0,\dots,n\}$、ω，$\varepsilon \in S_r$ 以及矩阵 $A \in \mathbb{R}^{n\times n}$，定义 $A(\omega|\varepsilon)$ 为删除指标出现在序列 ω 中的行，以及指标出现在 ε 中的列后所构成的矩阵。定义 $A[\omega|\varepsilon]$ 为删除指标未出现在 ω 中的行，以及指标未出现在 ε 中的列后所得到的矩阵。

引理 B.1（柯西-比内公式，[18] 中的方程 (1)）。令 A，$B \in \mathbb{R}^{n,\,n}$。我们有

$$\det(A+B) = \sum_{r=0}^{n} \sum_{\omega,\varepsilon\in S_r} (-1)^{s(\omega)+s(\varepsilon)} \det(A[\omega|\varepsilon]) \det(B(\omega|\varepsilon)).$$

引理 B.2（子式的柯西-施瓦茨不等式）。设 $A \in \mathbb{R}^{n\times n}$ 为对称半正定矩阵，且 i，j 是 $\{1,\dots,n\}$ 中互不相同的整数。我们有

$$\det(A(\omega\cup\{i\}|\omega\cup\{i\})) \det(A(\omega\cup\{j\}|\omega\cup\{j\})) \geq \det(A(\omega\cup\{i\}|\omega\cup\{j\}))^2.$$

证明。由于 A 是对称半正定矩阵，因此 $A(\omega|\omega)$ 也是如此。现在，所需的不等式可由以下矩阵这一事实推出

$$\operatorname{adj}(A(\omega|\omega)) = \lim_{\theta\to 0^+} \det(A(\omega|\omega) + \theta I)(A(\omega|\omega) + \theta I)^{-1}$$

是一个对称半正定矩阵。 □

引理 4.5 的证明。注意，核$(B^\top B)$ =核(B)，因此

$$\mathrm{Range}(B^\top B) = \mathrm{Ker}(B^\top B)^\perp = \mathrm{Ker}(B)^\perp = \mathrm{Range}(B^\top).$$

记矩阵 $B^\top B$为 $\mathcal{B}$，并记矩阵 $\mathcal{B}$的列为 B_1、.....、 B_n。只需证明对每个 $j \in \{1, \ldots, n\}$，命题对 $\bar{m} = B_j$成立。我们的目标是证明$[(\mathcal{B} + D_V^{-1})^{-1} B_j]_i$有界。

根据克拉默法则，我们有

$$[(\mathcal{B} + D_V^{-1})^{-1} B_j]_i = \frac{\det(E_i + D_{V,i}^{-1})}{\det(\mathcal{B} + D_V^{-1})}, \tag{44}$$

其中，矩阵 E_i 由将 $\mathcal{B}$ 的第 i 列替换为 B_j 得到；$D_{V,\ i}$ 则由将 D_V中的第 i个对角元素替换为 0 得到。

将 $E_i + D_{V\,i}^{-1}$ 的第 j 列从第 i 列中减去，我们得到

$$\det(E_i + D_{V,i}^{-1}) = \det(F_i + D_{V,i,j}^{-1}),$$

其中， F_i 是将矩阵 $\mathcal{B}$ 的第 i 列替换为 0 后得到的矩阵；$D_{V,\ i,\ j}$ 是将矩阵 D_V的第 i 列替换为 D_V的第 j 列的相反数后得到的矩阵。

根据引理 B.1，我们有

$$\begin{aligned}
\det(F_i + D_{V,i,j}^{-1}) &= \sum_r \sum_{\omega,\varepsilon \in S_r} (-1)^{s(\omega)+s(\varepsilon)} \det(F_i[\omega|\varepsilon]) \det(D_{V,i,j}^{-1}(\omega|\varepsilon)) \\
&= \sum_r \sum_{\substack{\omega \in S_{r-1} \\ \omega \not\ni i,j}} \det(F_i(\omega \cup \{j\}|\omega \cup \{i\}))\vartheta_\omega \prod_{k \in \omega \cup \{j\}} V_k^{-1} \\
&= \sum_r \sum_{\substack{\omega \in S_{r-1} \\ \omega \not\ni i,j}} \det(\mathcal{B}(\omega \cup \{j\}|\omega \cup \{i\}))\vartheta_\omega \prod_{k \in \omega \cup \{j\}} V_k^{-1},
\end{aligned}$$

其中 $\vartheta_\omega = \pm 1$。通过考察 i, j 是否属于 $\omega, \varepsilon \in S_r$ 的情况，可得前一个等式。

根据引理 B.2，我们有

$$\begin{aligned}
|\det(F_i + D_{V,i,j}^{-1})| \le &\max_{\varepsilon \not\ni i} \sqrt{\det(\mathcal{B}(\varepsilon \cup \{i\}|\varepsilon \cup \{i\}))} \\
&\times \sum_r \sum_{\substack{\omega \in S_{r-1} \\ \omega \not\ni i,j}} \sqrt{\det(\mathcal{B}(\omega \cup \{j\}|\omega \cup \{j\}))} \prod_{k \in \omega \cup \{j\}} V_k^{-1}.
\end{aligned} \tag{45}$$

另一方面，再次根据引理 B.1，我们有

$$\det(\mathcal{B} + D_V^{-1}) = \sum_r \sum_{\omega \in S_r} \det(\mathcal{B}(\omega|\omega)) \prod_{k \in \omega} V_k^{-1}. \tag{46}$$

最后需要指出的是，(45) 中的分子和 (46) 中的分母都是关于 $\{V_k^{-1}\}_k$ 的多项式，且次数相同。此外，(45) 中出现的所有多项式项也都出现在 (46) 中，且系数非零。进一步地，由于 $\mathcal{B} = B^\top B$ 是半正定的，(46) 中的系数是 $\mathcal{B}$ 的主子式的行列式，因此均为非负数。因此，分母没有正根（对于所有 $V \in (0, \infty)^n$，分母都严格为正）。因此，作为两个此类多项式之比的量 (44)，在 $V \in (0, \infty)^n$ 上一致有界。

□

引理 4.8 的证明。由于 K 可逆,

$$\mathrm{Null}(B)=\{x: AK^{-1}x=0\}=K\ \mathrm{Null}(A)=\mathrm{Range}(KU_A).$$

设 $W:=KU_A$。则 W 满列秩，且 值域(W) $=$零空间(B)。到 值域(W) 上的欧几里得正交投影算子为 $P_{\mathrm{Range}(W)}=W(W^\top W)^{-1}W^\top$。因此

$$P_{\mathrm{Null}(B)}=P_{\mathrm{Range}(W)}=KU_A\,(U_A^\top K^\top KU_A)^{-1}U_A^\top K^\top=KU_A\,G^{-1}U_A^\top K^\top.$$

与 $\mathbf{1}$ 取二次型，得到

$$\mathbf{1}^\top P_{\mathrm{Null}(B)}\mathbf{1}=\mathbf{1}^\top KU_A\,G^{-1}U_A^\top K^\top\mathbf{1}=z^\top G^{-1}z,$$

这就证明了 (33)。如果 $\bar m=\rho B^\top(Y-X\beta+\mu Bh)+\mu\mathbf{1}$，则 $P_{\mathrm{Null}(B)}B^\top=0$ 意味着 $\bar m_0=P_{\mathrm{Null}(B)}\bar m=\mu P_{\mathrm{Null}(B)}\mathbf{1}$，(34) 便立即成立。 □

引理 4.9 的证明。固定 $q\in(0,\frac12]$。为简便起见，记 $\tilde V:=\tilde V_i$ 和 $M:=M_i$。给定 M 的条件下，我们有

$$\tilde V\mid M\sim\mathrm{GIG}(\lambda,\psi,\chi),\qquad \lambda:=p-\tfrac12,\ \psi:=\tilde a>0,\ \chi:=b+\left(\tfrac{M}{\sigma}\right)^2\geq b>0.$$

设 $x:=\sqrt{\psi\chi}$。广义逆高斯分布的一个标准矩恒等式（当$\psi>0$且 $\chi>0$时，对所有实数 r 均成立）为

$$\mathbb{E}\Big[\tilde V^{\,r}\mid M\Big]=\left(\frac{\chi}{\psi}\right)^{r/2}\frac{K_{\lambda+r}(x)}{K_\lambda(x)},\qquad r\in\mathbb{R},\tag{47}$$

其中 $K_\nu(\cdot)$ 是第二类修正贝塞尔函数。

将 $r=-q$ 代入（47）得

$$\mathbb{E}\Big[\tilde V^{-q}\mid M\Big]=\left(\frac{\chi}{\psi}\right)^{-q/2}\frac{K_{\lambda-q}(x)}{K_\lambda(x)}.$$

步骤 1：对 M一致地控制前置因子。由于 $\chi\geq b$和 $q>0$，映射 $\chi\mapsto\chi^{-q/2}$单调递减，因此

$$\left(\frac{\chi}{\psi}\right)^{-q/2}\leq\left(\frac{b}{\psi}\right)^{-q/2}.$$

步骤 2：对 M一致地控制贝塞尔函数比值。注意，$x=\sqrt{\psi\chi}\geq\sqrt{\psi b}=:x_0>0$。定义连续正函数

$$R(x):=\frac{K_{\lambda-q}(x)}{K_\lambda(x)},\qquad x>0.$$

由于对所有 $x>0$ 和任意 $\nu\in\mathbb{R}$，$K_\nu(x)>0$，因此 R在$(0,\infty)$上定义良好且连续。此外，利用标准的大 x 渐近式 $K_\nu(x)\sim\sqrt{\pi/(2x)}e^{-x}$，当 $x\to\infty$时有 $R(x)\to1$。因此，R 在 $[x_0,\infty)$ 上有界：事实上，取 $X>x_0$ 使得对所有 $x\geq X$，$R(x)\leq2$，并令

$$C_K:=\max\Big\{\sup_{x\in[x_0,X]}R(x),\ 2\Big\}<\infty.$$

因此，对于所有 $M\in\mathbb{R}$,

$$\frac{K_{\lambda-q}(x)}{K_\lambda(x)}\leq C_K.$$

结合步骤 1–2 可得

$$\mathbb{E}\Big[\tilde{V}^{-q} \mid M\Big] \le \left(\frac{b}{\psi}\right)^{-q/2} C_K := L_-(p,\tilde{a},b,q) < \infty,$$

对 $M \in \mathbb{R}$ 一致成立。对 $\sum_{i=1}^{n} \mathbb{E}[\tilde{V}_i^{-q} \mid V]$ 的结论可通过对该界求和得到。

□

~ ~引理 4.10 的证明。令 $V := V_i$，且 $M := M_i$。与前面的证明一样，

$$\tilde{V} \mid M \sim \mathrm{GIG}(\lambda,\psi,\chi), \quad \lambda := p - \tfrac{1}{2},\ \psi := \tilde{a} > 0,\ \chi := b + \left(\tfrac{M}{\sigma}\right)^2,\ x := \sqrt{\psi\chi}.$$

将 $r = 1$ 代入 GIG 矩公式 (47)，得到

$$\mathbb{E}[\tilde{V} \mid M] = \left(\frac{\chi}{\psi}\right)^{1/2} \frac{K_{\lambda+1}(x)}{K_\lambda(x)} = \frac{\sqrt{\chi}}{\sqrt{\psi}} \cdot \frac{K_{\lambda+1}(x)}{K_\lambda(x)}.$$

我们使用修正贝塞尔函数的一个标准不等式（它可由该比值的连分数分析 / Riccati 分析推出；例如参见特殊函数文献中的 Amos 型界）：存在一个常数 $C_R(\lambda) < \infty$，使得对所有 $x > 0$,

$$\frac{K_{\lambda+1}(x)}{K_\lambda(x)} \le 1 + \frac{C_R(\lambda)}{x}. \tag{48}$$

（例如，可以取 $C_R(\lambda) = 2|\lambda| + 1$。）

利用 (48) 和 $x = \sqrt{\psi\chi}$，我们得到

$$\mathbb{E}[\tilde{V} \mid M] \le \frac{\sqrt{\chi}}{\sqrt{\psi}} + \frac{\sqrt{\chi}}{\sqrt{\psi}} \cdot \frac{C_R(\lambda)}{\sqrt{\psi\chi}} = \frac{\sqrt{\chi}}{\sqrt{\psi}} + \frac{C_R(\lambda)}{\psi}.$$

由于 $\chi = b + (M/\sigma)^2$,

$$\sqrt{\chi} \le \sqrt{b} + \left|\frac{M}{\sigma}\right|.$$

因此

$$\mathbb{E}[\tilde{V} \mid M] \le \frac{1}{\sqrt{\psi}} \left|\frac{M}{\sigma}\right| + \frac{\sqrt{b}}{\sqrt{\psi}} + \frac{C_R(\lambda)}{\psi}.$$

回顾 $\psi = \tilde{a}$ 和 $\lambda = p - \frac{1}{2}$，这恰好是

$$\mathbb{E}[\tilde{V}_i \mid M_i] \le \frac{1}{\sqrt{\tilde{a}}} \left|\frac{M_i}{\sigma}\right| + \frac{\sqrt{b}}{\sqrt{\tilde{a}}} + \frac{C_{\mathrm{lin}}(p,\tilde{a})}{\tilde{a}},$$

其中 $C_{\mathrm{lin}}(p,\ \tilde{a}) := C_R(p - \frac{1}{2})$。

□

引理 4.13 的证明。证明沿用了 [21] 中的类似论证。令

$$\tilde{a} := a + \frac{\mu^2}{\sigma^2}, \qquad \tilde{b} := \frac{M_i^2}{\sigma^2}, \qquad x := \sqrt{\tilde{a}\tilde{b}} = \frac{\sqrt{\tilde{a}}\,|M_i|}{\sigma}.$$

~下面我们假设 $b > 0$（即 $M_i \neq 0$）；此时该界在 $\pi(M_i|V,Y)\ -a.s.$ 内成立。利用广义逆高斯分布的标准矩恒等式,

$$E\left[V_i^{-\frac{\delta(p)}{2}} \mid M_i\right] = \int_0^\infty v^{-\frac{\delta(p)}{2}} \left(\frac{\sqrt{\tilde{a}}}{\sqrt{\tilde{b}}}\right)^{p-\frac{1}{2}} v^{(p-\frac{1}{2})-1}$$

$$\times \exp\left\{-\frac{1}{2}\left(\tilde{a}v+\frac{\tilde{b}}{v}\right)\right\}\frac{1}{2K_{p-\frac{1}{2}}\left(\sqrt{\tilde{a}\tilde{b}}\right)}\,dv$$

$$=\left(\frac{\sqrt{\tilde{a}}}{\sqrt{\tilde{b}}}\right)^{\frac{\delta(p)}{2}}\frac{K_{p-\frac{1}{2}-\frac{\delta(p)}{2}}\left(\sqrt{\tilde{a}\tilde{b}}\right)}{K_{p-\frac{1}{2}}\left(\sqrt{\tilde{a}\tilde{b}}\right)}.$$

我们考虑三种情况：$p<\frac{1}{2}$、$p>\frac{1}{2}$和 $p=\frac{1}{2}$。

情形 *1*：$p<\frac{1}{2}$。由 $K_\nu(x)=K_{-\nu}(x)$,

$$E\left[V_i^{-\frac{\delta(p)}{2}}\mid M_i\right]=\left(\frac{\sqrt{\tilde{a}}}{\sqrt{\tilde{b}}}\right)^{\frac{\delta(p)}{2}}\frac{K_{\frac{1}{2}-p+\frac{\delta(p)}{2}}\left(\sqrt{\tilde{a}\tilde{b}}\right)}{K_{\frac{1}{2}-p}\left(\sqrt{\tilde{a}\tilde{b}}\right)}.$$

应用命题 A.1，取 $\nu_1=\frac{1}{2}+\frac{\delta(p)}{2}-p$、$\nu_2=\frac{1}{2}-p$ 和 $x=\sqrt{\tilde{a}\tilde{b}}$：对于任意 $\epsilon^*>0$，存在 $\epsilon=\epsilon(\epsilon^*,\ p)$，使得对于 $\bar{x}<\epsilon$,

$$\frac{K_{\frac{1}{2}-p+\frac{\delta(p)}{2}}(x)}{K_{\frac{1}{2}-p}(x)}\le(1+\epsilon^*)\frac{\Gamma\left(\frac{1}{2}+\frac{\delta(p)}{2}-p\right)2^{\frac{\delta(p)}{2}}}{\Gamma\left(\frac{1}{2}-p\right)x^{\frac{\delta(p)}{2}}}.$$

应用命题 A.3（相同的 $\nu=\frac{1}{2}-p$）来界定当 $x\ge\epsilon$ 时的比值：

$$\frac{K_{\frac{1}{2}-p+\frac{\delta(p)}{2}}(x)}{K_{\frac{1}{2}-p}(x)}\le\frac{(2-2p)^{\frac{\delta(p)}{2}}}{x^{\frac{\delta(p)}{2}}}+1.$$

结合以上结果，并利用 $x^{-\frac{\delta(p)}{2}}=(\tilde{a}\tilde{b})^{-\frac{\delta(p)}{4}}$，我们得到

$$E\left[V_i^{-\frac{\delta(p)}{2}}\mid M_i\right]\le(1+\epsilon^*)\frac{\Gamma\left(\frac{1}{2}+\frac{\delta(p)}{2}-p\right)2^{\frac{\delta(p)}{2}}}{\Gamma\left(\frac{1}{2}-p\right)}\tilde{b}^{-\frac{\delta(p)}{2}}+\left(\frac{\tilde{a}}{\epsilon}\right)^{\frac{\delta(p)}{2}}\left(\frac{(2-2p)^{\frac{\delta(p)}{2}}}{\epsilon^{\frac{\delta(p)}{2}}}+1\right)$$

$$=(1+\epsilon^*)\frac{\Gamma\left(\frac{1}{2}+\frac{\delta(p)}{2}-p\right)2^{\frac{\delta(p)}{2}}}{\Gamma\left(\frac{1}{2}-p\right)}\left|\frac{M_i}{\sigma}\right|^{-\delta(p)}+C_{2,1}(p,\tilde{a}),$$

对于一个显式有限的 $C_{2,\ 1}(p,\ \tilde{a})$。

此外,

$$(1+\epsilon^*)\frac{\Gamma(\frac{1}{2}+\frac{\delta(p)}{2}-p)2^{\frac{\delta(p)}{2}}}{\Gamma(\frac{1}{2}-p)}\kappa(p)=(1+\epsilon^*)\frac{\left[\Gamma(\frac{1-p}{2})\right]^2}{\Gamma(\frac{1}{2}-p)\Gamma(\frac{1}{2})}\le(1+\epsilon^*)\frac{1-2p}{1-2p+\frac{p^2}{2}},$$

其中，最后一个不等式由 [15] 得出。因此，取 $\epsilon^*<2^{p^2}(1-2p)$，使得 $C_1(p)\kappa(p)<1$ 在此情形下成立。

情形 *2*：$p>\frac{1}{2}$。由于 $\delta(p)=$最小值$\{\frac{1}{2},2p-1\}$，我们有 $\delta(p)\le 2p-1$，因此

$$p-\frac{1}{2}-\frac{\delta(p)}{2}\ge p-\frac{1}{2}-\frac{2p-1}{2}=0,\qquad p-\frac{1}{2}-\frac{\delta(p)}{2}\le p-\frac{1}{2}.$$

因此 $0\le p-\frac{1}{2}-\frac{\delta(p)}{2}\le p-\frac{1}{2}$。利用 $K_{-\nu}(x)=K_\nu(x)$，并且由于 $\nu\mapsto K_\nu(x)$ 在 $\nu\ge 0$ 上对每个 $x>0$ 都是递增的,

$$\frac{K_{p-\frac{1}{2}-\frac{\delta(p)}{2}}(x)}{K_{p-\frac{1}{2}}(x)}\le 1.$$

因此,

$$E\left[V_i^{-\frac{\delta(p)}{2}} \mid M_i\right] \le \left(\frac{\sqrt{\tilde{a}}}{\sqrt{\tilde{b}}}\right)^{\frac{\delta(p)}{2}} \le \frac{1}{2\kappa(p)}\,\tilde{b}^{-\frac{\delta(p)}{2}} + \frac{\kappa(p)}{2}\,\tilde{a}^{\frac{\delta(p)}{2}} \quad \text{(Young)}$$
$$= \frac{1}{2\kappa(p)}\left|\frac{M_i}{\sigma}\right|^{-\delta(p)} + \frac{\kappa(p)}{2}\,\tilde{a}^{\frac{\delta(p)}{2}},$$

因此，这里有 $C_1(p) = 1/(2\kappa(p))$，并且 $C_1(p)\kappa(p) = 1/2 < 1$。

情形 *3*: $p = \frac{1}{2}$。这里 $\delta(p) = \frac{1}{2}$。根据命题 A.2 和 A.3，存在 $\epsilon' > 0$，使得对于 $x = \sqrt{\tilde{a}\tilde{b}}$,

$$\frac{K_{\frac{\delta(p)}{2}}(x)}{K_0(x)} \le \frac{2^{\frac{1}{4}}\sqrt{\pi}}{2\Gamma(\frac{1}{4})\,x^{\frac{\delta(p)}{2}}}\,I_{(0<x<\epsilon')} + \left(\frac{1}{x^{\frac{\delta(p)}{2}}} + 1\right) I_{(x\ge\epsilon')}.$$

因此

$$E\left[V_i^{-\frac{\delta(p)}{2}} \mid M_i\right] = \left(\frac{\sqrt{\tilde{a}}}{\sqrt{\tilde{b}}}\right)^{\frac{\delta(p)}{2}} \frac{K_{\frac{\delta(p)}{2}}(x)}{K_0(x)}$$
$$\le \frac{2^{\frac{1}{4}}\sqrt{\pi}}{2\Gamma(\frac{1}{4})}\,\tilde{b}^{-\frac{\delta(p)}{2}} + C_{2,3}(p,\tilde{a}) = \frac{2^{\frac{1}{4}}\sqrt{\pi}}{2\Gamma(\frac{1}{4})}\left|\frac{M_i}{\sigma}\right|^{-\delta(p)} + C_{2,3}(p,\tilde{a}),$$

可得

$$\frac{2^{\frac{1}{4}}\sqrt{\pi}}{2\Gamma(\frac{1}{4})}\,\kappa\big(\tfrac{1}{2}\big) = \frac{1}{2} < 1.$$

结论。综合这三种情形可得

$$E\left[V_i^{-\frac{\delta(p)}{2}} \mid M_i\right] \le C_1(p)\left|\frac{M_i}{\sigma}\right|^{-\delta(p)} + C_2(p,\tilde{a}),$$

with

$$C_1(p) = (1+\epsilon^*)\frac{\Gamma(\frac{1}{2} + \frac{\delta(p)}{2} - p)2^{\frac{\delta(p)}{2}}}{\Gamma(\frac{1}{2} - p)}\,I_{(p<\frac{1}{2})} + \frac{1}{2\kappa(p)}\,I_{(p\ge\frac{1}{2})},$$

以及一个有限的 $C_2(p,\ \tilde{a})$，它汇集了各情形中的相应常数。此外，$C_1(p)\kappa(p) < 1$ 对所有 $p > 0$ 均成立，这由上述论证可知。 □

引理 4.14 的证明。根据 ℓ_∞-对偶性，我们有

$$\|\bar{Q}(V)^{-1}u\|_1 = \sup_{\|s\|_\infty \le 1} s^\top \bar{Q}(V)^{-1}u.$$

For 对于任意这样的 s，在 $\bar{Q}$（V）$^{-1}$-内积下应用柯西–施瓦茨不等式，得到 s

$$s^\top \bar{Q}(V)^{-1}u \le \sqrt{s^\top \bar{Q}(V)^{-1}s}\,\sqrt{u^\top \bar{Q}(V)^{-1}u}.$$

由于 $\rho B^\top B \succeq 0$，我们有 $\bar{Q}$（V）$\succeq D_V^{-1}$。这意味着 $\bar{Q}$（V）$^{-1} \preceq D_V$。因此,

$$s^\top \bar{Q}(V)^{-1}s \le s^\top D_V s = \sum_{i=1}^n V_i s_i^2 \le \sum_{i=1}^n V_i,$$

最后一个不等式成立是因为 $|s_i| \le 1$。对 $\|s\|_\infty \le 1$ 取上确界即可得到 (35)。

□

附录 C：显式梯度推导

本附录详细推导第 1.2 节所述 LLnGM 模型的联合对数似然函数的梯度估计量。我们明确比较中心化与非中心化参数化下的公式，以突出前者固有的可积性问题。

C.1. 参数化与分布。 遵循 [34] 和 [24], , 为确保方差过程属于在卷积下封闭的一类分布，我们只能考虑广义双曲（GH）分布的两种特殊情形：正态逆高斯（NIG）分布和广义非对称拉普拉斯（GAL）分布。为避免对式（3）中的 B 过度参数化，我们采用 [2]中的参数化方式。具体而言，令 $\nu > 0$为方差分量参数：对于 NIG 噪声， $V_i \sim$ 逆高斯分布$(\nu,\ \nu h_i^2)$；对于 GAL 噪声，$V_i \sim$ 伽马分布$(h_i\nu,\ \nu)$。

C.2. 中心化参数化。 令 $Y = (y_1,\ \ldots,\ y_m)^\top$、 $W = (w_1,\ \ldots,\ w_n)^\top$，以及 ζ 表示核参数向量，并令 θ 表示完整参数向量（包括ζ、 μ、 σ等）。联合对数似然为

$$\log p_\theta(Y, W, V) = \log p_\theta(Y \mid W) + \log p_\theta(W \mid V) + \log p_\theta(V),$$

其中各个分量为：

$$\log p_\theta(Y \mid W) = -\frac{m}{2}\log(2\pi) - m\log\sigma_\epsilon - \frac{1}{2\sigma_\epsilon^2}\|Y - X\beta - AW\|^2,$$

$$\log p_\theta(W \mid V) = -\frac{n}{2}\log(2\pi) - n\log\sigma + \log|K| - \frac{1}{2}\sum_{i=1}^{n}\log V_i$$
$$- \frac{1}{2\sigma^2}(KW - \mu(V-h))^\top D_V^{-1}(KW - \mu(V-h)),$$

其中 $D_V = \mathrm{diag}(V)$。项 $\log p_\theta(V)$ 取决于混合分布：

• 对于 $V_i \sim$ 逆高斯分布$(\nu, \nu h_i^2)$:

$$\log p_\theta(V) = \sum_{i=1}^{n}\left(\frac{1}{2}\log\left(\frac{\nu h_i^2}{2\pi V_i^3}\right) - \frac{\nu V_i^2 + \nu h_i^2}{2V_i} + \nu h_i\right).$$

• 对于 $V_i \sim$ 伽马分布$(h_i\nu, \nu)$:

$$\log p_\theta(V) = \sum_{i=1}^{n}\left(h_i\nu\log(\nu) - \log\Gamma(h_i\nu) + (h_i\nu - 1)\log V_i - \nu V_i\right).$$

现在，我们推导联合对数似然 $\mathcal{L} = \log p_\theta(Y, W, V)$ 关于各个参数的梯度。

回归系数 β。 关于回归系数的梯度为

$$\nabla_\beta \mathcal{L} = \frac{1}{\sigma_\epsilon^2} X^\top (Y - X\beta - AW).$$

*N*噪声方差 σ_ϵ。 对观测噪声方差求导得 s

$$\partial_{\sigma_\epsilon}\mathcal{L} = -\frac{m}{\sigma_\epsilon} + \frac{1}{\sigma_\epsilon^3}\|Y - X\beta - AW\|^2.$$

*P*过程尺度 σ。 展开二次项 $((KW)_i - \mu V_i + \mu h_i)^2/V_i$，得到 :

$$\begin{aligned}\partial_\sigma \mathcal{L} &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}\frac{((KW)_i - \mu(V_i - h_i))^2}{V_i} \\ &= -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}\left[\frac{(KW)_i^2}{V_i} + \mu^2 V_i + \frac{\mu^2 h_i^2}{V_i} - 2\mu(KW)_i + \frac{2\mu h_i (KW)_i}{V_i} - 2\mu^2 h_i\right].\end{aligned}$$

注意，该表达式包含显式项 $+\mu^2 h_i^2/V_i$。虽然似然项可能会抑制 $V_i \to 0$ 处的奇异性，但要验证该得分函数的后验可积性在技术上颇为棘手，尤其是在形状参数较小（$h_i\nu \le 1$）的 GAL 情形下，此时先验逆矩 $\mathbb{E}[V_i^{-1}]$ 不存在。

漂移参数 μ。 关于漂移参数的梯度为

$$\begin{aligned}\partial_\mu \mathcal{L} &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\frac{((KW)_i - \mu V_i + \mu h_i)(V_i - h_i)}{V_i} \\ &= \frac{1}{\sigma^2}\sum_{i=1}^{n}\left[(KW)_i - \mu V_i + 2\mu h_i - \frac{h_i (KW)_i}{V_i} - \frac{\mu h_i^2}{V_i}\right].\end{aligned}$$

与 $\partial_\sigma \mathcal{L}$ 类似，含有按 $1/V_i$ 缩放的项（例如 $-\mu h_i^2/V_i$）会给使用标准漂移条件建立 L^1 收敛保证带来挑战。

核参数 ζ_j。 对于每个核参数 ζ_j，其梯度为

$$\partial_{\zeta_j}\mathcal{L} = \operatorname{tr}(K^{-1}\partial_{\zeta_j}K) - \frac{1}{\sigma^2}(KW - \mu(V-h))^\top D_V^{-1}(\partial_{\zeta_j}K)W.$$

混合参数 ν。 关于混合参数的梯度取决于分布的选择:

$$\partial_\nu \mathcal{L} = \begin{cases}\sum_{i=1}^{n}\left(\frac{1}{2\nu} + h_i - \frac{V_i}{2} - \frac{h_i^2}{2V_i}\right) & \text{for NIG,} \\ \sum_{i=1}^{n}\left(h_i\log(\nu) + h_i - h_i\psi(h_i\nu) + h_i\log V_i - V_i\right) & \text{for GAL,}\end{cases} \tag{49}$$

其中 $\psi(\cdot)$ 是双伽马函数。

C.3. 非中心化参数化。在非中心化参数化中，我们定义 $M = KW + \mu h$，使得 $W = K^{-1}(M - \mu h)$。似然分量变为:

$$\log p_\theta(Y \mid M) = -\frac{m}{2}\log(2\pi) - m\log\sigma_\epsilon - \frac{1}{2\sigma_\epsilon^2}\|Y - X\beta - AK^{-1}(M - \mu h)\|^2,$$

$$\log p_\theta(M \mid V) = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2}\sum_{i=1}^{n}\log V_i - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\frac{(M_i - \mu V_i)^2}{V_i}.$$

该 formulation 的梯度可通过类似方式推导得到。

回归系数 β。

$$\nabla_\beta \log p_\theta = \frac{1}{\sigma_\epsilon^2}X^\top\left(Y - X\beta - AK^{-1}(M - \mu h)\right).$$

噪声方差 σ_ϵ。

$$\partial_{\sigma_\epsilon}\log p_\theta = -\frac{m}{\sigma_\epsilon} + \frac{1}{\sigma_\epsilon^3}\|Y - X\beta - AK^{-1}(M - \mu h)\|^2.$$

过程尺度 σ。

$$\partial_\sigma \log p_\theta = -\frac{n}{\sigma} + \frac{1}{\sigma^3}\sum_{i=1}^{n}\frac{(M_i-\mu V_i)^2}{V_i}.$$

此处，项$(M_i-\mu V_i)^2/V_i$是一个自归一化二次型。在高斯层$M \mid V$下，它等于 $\sigma^2 Z_i^2$，其中 $Z_i \sim N(0,1)$；更一般地，在使用条件高斯分布 $M \mid (V,Y)$时，它在$V_i \to 0$ 附近仍表现良好。这种结构使非中心化参数化下的矩界更加清晰。

漂移参数 μ。 对对数 $p(Y|M)$ 和对数 $p(M|V)$同时应用链式法则：

$$\partial_\mu \log p_\theta = \frac{1}{\sigma^2}\sum_{i=1}^{n}(M_i-\mu V_i) + \frac{1}{\sigma_\epsilon^2}(AK^{-1}h)^\top\left(Y - X\beta - AK^{-1}(M-\mu h)\right).$$

核参数 ζ_j。

$$\partial_{\zeta_j} \log p_\theta = \frac{1}{\sigma_\epsilon^2}\left(Y - X\beta - AK^{-1}(M-\mu h)\right)^\top AK^{-1}(\partial_{\zeta_j}K)K^{-1}(M-\mu h).$$

ν 的梯度与中心化情形相同。蒙特卡罗梯度估计量和 Rao–Blackwell 化梯度估计量可通过将(W,V)（中心化）或(M,V)（非中心化）替换为吉布斯样本得到；对于 Rao–Blackwell 化版本，则分别将相应的潜在高斯变量替换为在 $W \mid (V,Y)$ 或 $M \mid (V,Y)$下的条件期望。

参考文献

[1] ABRAMOWITZ, M. 和 STEGUN, I.（1965）。数学函数手册：公式、图表与数学表。*Dover*高等数学丛书。Dover出版社，纽约。[2] ASAR, O.、BOLIN, D.、DIGGLE, P. J. 和 WALLIN, J.（2020）。用于非高斯连续重复测量数据的线性混合效应模型。皇家统计学会期刊：*C*辑（应用统计）**69** 1015-1065。[3] BAKKA, H.、RUE, H.、FUGLSTAD, G.-A.、RIEBLER, A.、BOLIN, D.、ILLIAN, J.、KRAINSKI, E.、SIMP-SON, D. 和 LINDGREN, F.（2018）。使用R-INLA进行空间建模：综述。*WIREs*计算统计 **10** 1443 10 1002 1443 e。https://doi.org/。/wics。

[4] BARNDORFF-NIELSEN, O.（1978）。双曲分布与双曲线上的分布。*Scandinavian Journal of Statistics***5**。[5] BHATTACHARYA, A.、PATI, D.、PILLAI, N. S. 和 DUNSON, D. B.（2015）。用于最优收缩的狄利克雷–拉普拉斯先验。*Journal of the American Statistical Association* **110** 1479–1490。[6] BHATTACHARYA, S.、KHARE, K. 和 PAL, S.（2022）。马蹄先验及其正则化变体的吉布斯采样器的几何遍历性。*Electronic Journal of Statistics* **16** 1–57。https://doi.org/10.1 214/21-EJS1932[7] BOLIN, D.（2014）。由非高斯噪声驱动的空间Matérn场。*Scandinavian Journal ofStatistics* **41** 557 579 10 1111 12046 -。https://doi.org/。/sjos。BOLIN, D. 和 WALLIN, J.（2020）。多元G型Matérn随机偏微分方程随机场。皇家统计学会期刊：*B*辑（统计方法学）**82** 215– 239 10 1111 12351 。多元[8] 模型。https://doi.org/。/rssb。BOTTOU, L.（1998）。在线算法与随机逼近。剑桥大学出版社。[9] CABRAL, R.、BOLIN, D. 和 RUE, H.（2024）。使用变分贝叶斯和拉普拉斯近似拟合潜变量非高斯模型。*Journal of t he American Statistical Association* **119** 1–13。https: //doi.org/ 10 1080 01621459 2023 2296704 . / . 。CARVALHO, C. M.、POLSON, N. G. 和 SCOTT, J. G.（2010）。用于稀疏信号的马蹄估计量。*Biometrika* **97** 465 480 10 1093 017 –。https://doi.org/。/biomet/asq [11]

[12] DHULL, M. S. 和 KUMAR, A.（2021）。使用EM算法的正态逆高斯自回归模型。国际工程科学与应用数学进展杂志**13** 139 147 10 1007 12572 021 00303 –。https://doi.org/。/s - - -y

[13] DOUC, R.、MOULINES, E. 和 STOFFER, D.（2013）。非线性时间序列：理论、方法及R示例应用。Chapman and Hall/CRC，纽约。https://doi.org/

10 1201 16331 . /b FONG, Y.、RUE,

[14] 2010
H. 和 WAKEFIELD, J.（）。广义线性混合模型的贝叶斯推断。Biostatistics（英国牛津）

11 397 412 10 1093 053 –。https://doi.org/。/

[15] 1956
biostatistics/kxp GURLAND, J.（）。伽马函数满足的一个不等式。Skand. Aktuarietidskr. 39 171–172。
https://doi.org/ 10 1080 03461238 1956 10414947 95293 . / . 。MR KHARE, K. 和 HOBERT, J.

[16] 2013
P.（）。贝叶斯套索的几何遍历性。Electronic Journal of Statistics 7 2150–2163。[17] KUSHNER, H. J. 和 YIN, G. G.（2003）。随机逼近与递归算法及其应用，第2版。随机建模与应用概率35。Springer，纽约州纽约。https: //doi.org/

10 1007 97441 . /b MARCUS, M.（）。和的行列式。The College Mathematics Journal 21 130–135。

[18] 1990
[19] MARTINO, S.、AKERKAR, R. 和 RUE, H.（2011）。生存模型的近似贝叶斯推断。Scandinavian Journal of Statistics 38 514–528。https://doi.org/10.1111/j.1467-9469.2010.00715.x [20] PACIOREK, C. 和 SCHERVISH, M.（2003）。用于高斯过程回归的非平稳协方差函数。载于 Advances in Neural Information Processing Systems 16。MIT出版社。[21] PAL, S. 和 KHARE, K.（2014）。贝叶斯收缩模型的几何遍历性。Electronic Journal of Statistics

8 10 1214 14 896 。https://doi.org/。/ -EJS PAL, S.、KHARE, K. 和 HOBERT, J.

[22] 2017
P.（）。使用广义双帕累托收缩先验进行贝叶斯推断的迹类马尔可夫链。Scandinavian Journal of Statistics 44
307-323。https: //doi.org/ 10 1111 12254 . /sjos. PARK, T. 和 CASELLA, G.（）。贝叶斯套索。

[23] 2008
Journal of the American Statistical Association 103 681–686。[24] PODGORSKI, K. 和 WALLIN, J.（2016）。广义双曲分布的卷积不变子类。Communications in Statistics - Theory and Methods 45 98 03 610926 103 10 1080 2013 821489 –。https://doi.org/。/ . .[25] QIN, Q.、HOBERT, J. P. 和 KHARE, K.（2019）。迹类马尔可夫算子谱隙的估计。Electronic Journal of Statistics

13 1790 1822 10 1214 19 1563 –。https://doi.org/。/ -

[26] 1951
EJS ROBBINS, H. 和 SUTTON, M.（）。一种随机逼近方法。Ann. Math. Statist. 22 400–407。[27] ROBERTS, G. 和 ROSENTHAL, J.（1997）。几何遍历性与混合马尔可夫链。Electronic Communications in Probability

2 13 25 10 1214 2 981 –。https://doi.org/。/ECP.v -

[28] 2004
ROBERTS, G. O. 和 ROSENTHAL, J. S.（）。一般状态空间马尔可夫链与马尔可夫链蒙特卡罗算法。Probability Surveys **1** 20 71 10 1214 154957804100000024 –。https://doi.org/。/

[29] 1994
ROBERTS, G. O. 和 SMITH, A. F. M.（）。吉布斯采样器和Metropolis-Hastings算法收敛的简单条件。Stochastic Processes and their Applications 49 207–216。[30] ROSENTHAL, J. S.（1995）。马尔可夫链蒙特卡罗的小化条件与收敛速率。Journal of the American Statistical Association

90 558 566 10 2307 2291067 –。https://doi.org/。/

[31]
RUE, H.、RIEBLER, A.、SØRBYE, S. H.、ILLIAN, J. B.、SIMPSON, D. P. 和 LINDGREN, F. K.（2017）。使用INLA进行贝叶斯计算：综述。Annual Review of Statistics and Its Application 4 395–

421 10 1146 031016 085240 。https://doi.org/。/annurev-statistics- -

[32] 1994
TIERNEY, L.（）。用于探索后验分布的马尔可夫链。The Annals of Statistics 22 1701 1728–。[33] WALDER, A. 和 HANKS, E. M.（2020）。利用拉普拉斯移动平均随机场对空间广义线性混合模型进行贝叶斯分析。Computational Statistics & Data Analysis 144 106861。https://doi.org/ 10 1016 2019 106861 . /j.csda. .

[34] 2015
WALLIN, J. 和 BOLIN, D.（）。使用非高斯Matérn场进行地统计建模：非高斯Matérn场。Scandinavian Journal of Statistics 42 872–890。https://doi.org/10.1111/sjos .12141[35] ZHANG, L.（2017）。贝叶斯收缩模型的迹类马尔可夫链，博士学位论文，佛罗里达大学。[36] ZHANG, L.、KHARE, K. 和 XING, Z.（2019）。正态-伽马贝叶斯收缩模型的迹类马尔可夫链。Electronic Journal of Statistics 13 166–207。https://doi.org/10.1214/18-EJS1491